(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number
WO 01/77380 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/GB01/01622

(22) International Filing Date: 10 April 2001 (10.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
0008801.3 10 April 2000 (10.04.2000) GB

(71) Applicant *(for all designated States except US)*: **MEDICAL RESEARCH COUNCIL** [GB/GB]; 20 Park Crescent, London W1N 4AL (GB).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: **WRIGHT, Alan** [GB/GB]; MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh EH4 2XU (GB).

(74) Agents: **HARDING, Charles, Thomas** et al.; D Young & Co., 21 New Fetter Lane, London EC4A 1DA (GB).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii)) for the following designations AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG)*

— *as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii)) for all designations*

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/77380 A2

(54) Title: SEQUENCES

(57) Abstract: Methods and means are described for diagnosing a disease or a predisposition to disease by genotyping a *RPGR* gene from an individual and identifying the presence of one or more disease causing mutation(s). Methods of treatment, kits for diagnostic purposes and an isolated *RPGR* gene including all or some of a range of mutations are also described as well as methods of isolating molecules that modulate a *RPGR* gene.

50

SEQUENCES

FIELD OF THE INVENTION

The present invention relates to novel sequences.

In particular, the present invention relates to novel amino acid sequences and novel nucleotide sequences encoding same.

The present invention also relates to compositions – such as pharmaceutical compositions and/or diagnostic compositions - containing or targeting one or more of said sequences.

The present invention also relates to assays utilising said sequences and methods of detecting the presence or absence of one or more of said sequences.

The present invention also relates to a method for determining mutation(s) in a gene, as well as means for using such a method in therapeutic applications.

In addition, the present invention relates to a kit for diagnosis for susceptibility or predisposition to a disease.

The present invention also relates to a method for the diagnosis of a disease or a predisposition to a disease by screening for the presence of mutation(s) in a gene.

The present invention further relates to directed treatment of such disease states.

BACKGROUND TO THE INVENTION

X-linked retinitis pigmentosa (XLRP) is a form of retinitis pigmentosa (RP). XLRP is clinically one of the most severe forms of RP, with onset in the first decade of life and severe visual impairment by the fourth decade.

XLRP affects 16-33% of all RP patients and genetic mapping studies suggested that about 75% of families mapped to chromosomal region Xp21.1. A gene was isolated in from Xp21.1 in 1996 which was found to be responsible for mutations in 15-20% of XLRP patients (Meindl et al., 1996), which was later confirmed by others.

The diagnosis of XLRP has major implications for families since a female carrier will have a 1 in 2 chance of having a son with severe disease. There is therefore considerable demand for an efficient diagnostic test for XLRP.

However, the diagnosis of XLRP is difficult since there are no clinical means of reliably distinguishing it from other forms of RP.

The present invention seeks to address this problem.

SUMMARY ASPECTS

The present invention is based on the novel finding that it is possible to reliably diagnose for the presence of, or a pre-disposition to, XLRP by identifying disease causing mutation(s) within a *RPGR* gene sequence.

Thus, the present invention relates to methods for *inter alia* identifying and/or diagnosing the presence or absence of one or more disease causing mutation(s) within a *RPGR* gene sequence. In particular, these methods relate to screens to determine the presence or absence of a disease causing mutation, such as single nucleotide mutation. The methods of the present invention may also be used to determine the relative position of multiple disease causing mutation(s) within a *RPGR* gene sequence in order to provide a set of disease causing mutation(s) or a haplotype for a *RPGR* gene in an individual. The identified disease causing mutation(s) may be used to diagnose a disease and/or predisposition to disease by correlating the identified disease causing mutation(s) with inherited genetic factors and/or phenotypic traits. The identified disease causing mutation(s) in a *RPGR* gene may be used as targets for the discovery of agents (such as

modulators) which may be effectively used to prevent or delay or treat a disease or a predisposition to a disease associated with these genetic variations.

DETAILED ASPECTS OF THE INVENTION

According to one aspect of the present invention, there is provided a method of diagnosis for a disease or a predisposition to a disease associated with a disease causing mutation in a *RPGR* gene; and wherein the method comprises: (i) genotyping a *RPGR* gene; and (ii) determining whether the genotype comprises a disease causing mutation.

In this embodiment, typically the *RPGR* gene is taken from an individual or is in a sample taken from an individual.

Typically the individual is a human.

According to another aspect of the present invention there is provided a kit for diagnosis of a disease or a predisposition to disease, wherein the kit comprises: (i) means for genotyping a *RPGR* gene; and (ii) reference means for determining whether the genotype comprises a disease causing mutation.

According to another aspect of the present invention there is provided a mutant *RPGR* gene, wherein said gene comprises one or more disease causing mutations.

According to another aspect of the present invention there is provided a nucleotide sequence capable of selectively hybridising to a mutant *RPGR* gene (and not the wild-type *RPGR* gene); wherein said gene comprises one or more disease causing mutations.

According to another aspect of the present invention there is provided a mutant RPGR protein encodable by said mutant gene.

The present invention also encompasses novel sequences, as well as variants, homologues, derivatives or fragments thereof. These sequences are presented as SEQ

ID No 1 and SEQ ID No. 2, and the series of sequences presented as SEQ ID No. 3. The present invention encompasses diagnostic methods for identifying said sequences, as well as kits comprising means for achieving same.

Other aspects of the present invention are presented in the accompanying claims and in the following description and drawings. These aspects are presented under separate section headings. However, it is to be understood that the teachings under each section are not necessarily limited to that particular section heading.

PREFERABLE ASPECTS

Preferably the or each disease causing mutation is located within ORF15 of the *RPGR* gene. ORF15 is presented as SEQ ID No. 1.

Preferably the or each disease causing mutation is located within a mutation hot spot of ORF15 of the *RPGR* gene. This mutation hot spot of ORF15 is presented as SEQ ID No. 2.

Preferably the disease causing mutation is one or more of the sequences presented under SEQ ID No. 3.

Preferably the diagnostic method is carried out using one or more PCR primers.

Preferably the PCR primer(s) is/are capable of selectively hybridising to some or all of the sequence presented as SEQ ID No. 1.

Preferably the PCR primer(s) is/are complementary to some or all of the sequence presented as SEQ ID No. 1.

Preferably the PCR primer(s) is/are capable of selectively hybridising to some or all of the sequence presented as SEQ ID No. 2.

Preferably the PCR primer(s) is/are complementary to some or all of the sequence presented as SEQ ID No. 2.

Preferably the genotyping (which may be a diagnostic method) is carried out using allelic specific primers.

Preferably the PCR primer(s) is/are capable of selectively hybridising to some or all of the sequences presented as SEQ ID No. 3.

Preferably the PCR primer(s) is/are complementary to some or all of the sequences presented as SEQ ID No. 3.

Preferably PCR techniques are used to genotype a nucleic acids comprising a *RPGR* gene or part thereof from an individual.

Preferably the results of genotyping of *RPGR* disease causing mutation(s) may be used to identify patients that are highly likely to suffer from certain disease state(s).

SURPRISING AND UNEXPECTED FINDINGS

The present invention demonstrates the surprising and unexpected findings that disease causing mutation(s) exist in the *RPGR* gene, which disease causing mutation(s) are accountable for a certain disease state.

The *RPGR* gene of the present invention (which is sometimes referred to as a mutant *RPGR* gene) is different to the wild type sequence.

Hence, some embodiments of the present invention are based on methods for identifying *RPGR* genes other than the wild type *RPGR* gene. The genes which are to be identified include variant or allelic *RPGR* genes with disease causing mutation(s). It is to be understood that these variant *RPGR* genes may be identified by reference to either the wild type *RPGR* gene or another reference/control sequence.

ADVANTAGES

The present invention is advantageous because it facilitates the genotyping of *RPGR* gene disease causing mutation(s) which in turn:

(i) provides for a more accurate diagnosis of a predisposition to a certain disease state. Thus, by genotyping the *RPGR* gene, an individual may be identified as being predisposed to a certain disease state.

(ii) allows for the identification of individuals who are predisposed to a certain disease state or who have an increased risk of contracting such a certain disease state. A suitable therapy may then be put in place to prevent or treat or delay the onset of these diseases.

(iii) helps to identify patients most likely to respond positively to treatment with certain classes of therapies or particular therapeutics.

(iv) allows for the selection of optimal clinical trial patient samples thereby reducing the size of a trial and/or decreasing the time of the clinical trial.

Other advantages are discussed and are made apparent in the following commentary.

RPGR

The present invention invention concerns disease causing mutation(s) in the *RPGR* gene. The gene comprising said disease causing mutation(s) is different to the wild-type gene – and thus may be termed a mutant. The mutation may be a single disease causing mutation or multiple disease causing mutation(s).

Background teaching on *RPGR* have been presented by Victor A McKusick in "Online Mendelain Inheritance in Man (OMIM)", John Hopkins University, Baltimore, MD (see www.ncbi.lm.nih.gov/Omim). For ease of reference, teachings from that source are now repeated below:

Falls and Cotterman (1948) described an X-linked form of choroidoretinal degeneration which is distinguished from other types by the presence in heterozygous women of a tapetal-like retinal reflex (a brilliant, scintillating, golden-hued, patchy appearance most striking around the macula) but no visual defect. See retinitis pigmentosa-2 for a phenotypically related entity. It had long been thought that there was probably more than one X-linked locus leading to a retinitis pigmentosa type of picture, and this was corroborated by the findings of linkage studies (see later).

In a large kindred segregating for X-linked recessive retinitis pigmentosa with metallic-sheen fundus reflex in heterozygotes, Nussbaum et al. (1985) found measurable linkage to DXS7 (maximum lod = 2.5 at theta = 0.125). This is the same RFLP as that shown to be tightly linked to other forms of X-linked retinitis pigmentosa (XLRP) (Bhattacharya et al., 1984). The 95% probability limits are such that these findings might indicate allelism of these clinically different forms of RP. Studies with other RFLPs placed this form of RP distal to DXS7 on Xp. Musarella et al. (1987) found close linkage of a form of X-linked RP and OTC with an anonymous DNA marker, 754, at Xp21 (interval = about 6 cM; lod = greater than 3.0). Chen et al. (1987) and Wirth et al. (1987, 1988) also found close linkage of one form of RP to OTC. Denton et al. (1988) did linkage studies in 3 large pedigrees segregating for the form of X-linked RP with the characteristic tapetal reflex in heterozygotes. Very close linkage to OTC was found (lod = 10.463 at theta = 0.01). Thus, the form of RP is probably that referred to here as RP3. It is also the locus presumably deleted in BB, the boy with RP, Duchenne muscular dystrophy, chronic granulomatous disease, and McLeod syndrome (Francke et al., 1985). Curtis and Blank (1989) studied a family in which a carrier female had an unusual tapetal reflex, the macula having 'a beaten metal appearance, with glistening patches.' The data supported the conclusion that retinitis pigmentosa with tapetal reflex is a separate entity. In another large kindred with X-linked retinitis pigmentosa and metallic sheen in the heterozygous carriers, Musarella et al. (1989) again found close linkage with Xp21 marker loci OTC and DXS206. By multipoint linkage analysis applying heterogeneity tests in 20 X-linked RP families, Musarella et al. (1990) concluded that the second X-linked RP locus may be located 8.5 cM proximal to DXS28 at Xp21.3. Chen et al. (1989) likewise found evidence of 2 distinct RP loci on Xp. In 1 family, they found the disease locus to be centromeric to DXS7, whereas in another family it was telomeric to DXS7. In 1 of 3 Swedish families, Dahl et al. (1991) demonstrated that the RP locus mapped to the same position as OTC and therefore represented RP3. In the other 2 families, linkage to OTC was excluded.

Fujita et al. (1996) analyzed 27 individuals with X-linked RP from a large American family of apparent Irish descent, using 17 polymorphic markers for linkage analysis. Segregation of XLRP with markers in Xp21.1 was consistent with the RP3 subtype. A recombination proximal to DXS1110 (between markers DXS8349 and M6) was found in 1 patient with RP3, placing the mutation locus outside the deletion breakpoint, located about 40 kb centromeric to DXS11110, of patient BB reported by Francke et al. (1985).

In a family with retinitis pigmentosa presumed to be RP3 because of linkage evidence, van Dorp et al. (1992) found that some affected males had recurrent respiratory infections as a result of a condition indistinguishable from the immotile cilia syndrome. They raised the possibility that previously observed ciliary abnormalities in XLRP patients may be associated specifically with the RP3 locus mutation. Infertility of the affected males was not a feature. Abnormalities of cilia have been reported in X-linked and autosomal types of RP, including Usher syndrome Arden and Fox, 1979; Fox et al., 1980; Hunter et al., 1988). Keith et al. (1991) described a large Australian family with extreme clinical variability in the hemizygotes: 1 member had typical rod-cone disease, 3 had the cone-rod pattern,

and 1 had macroscopic changes in the macular area only, but showed low potentials in the ERG. The locus for the disorder in this family was found to be distal to L128 at Xp21. From a study of reported case histories, Keith et al. (1991) concluded that clinical variability is a common feature of X-linked retinitis pigmentosa.

Meindl et al. (1996) isolated and sequenced cosmids from the region of the microdeletions in RP3 patients and used these cosmids to make exon predictions. They thus identified a gene, provisionally named RPGR (retinitis pigmentosa GTPase regulator), which gives rise to a ubiquitously expressed 29-kb transcript. The predicted RPGR protein has a series of tandemly arranged repeats characteristic of the highly conserved guanine nucleotide exchange factor, which regulates the GTPase RAN. Meindl et al. (1996) identified 8 potential asparagine-linked glycosylation sites along the N-terminal two-thirds of the predicted RPGR protein. They found that the C terminus of the protein contains a cluster of basic residues followed by a consensus isoprenylation site. They noted that confirmation of the isoprenylation of this site would establish a novel means of membrane anchorage for a GTPase regulator. Meindl et al. 1996) provided evidence that loss-of-function mutations within RPGR are responsible for RP3 type X-linked retinitis pigmentosa by identifying 2 small intragenic deletions and 2 nonsense and 3 missense mutations in highly conserved residues in unrelated patients with X-linked RP.

Ott et al. (1990) mapped the RP3 gene to a chromosome interval of less than 1000 kb between the DXS1110 marker and the OTC locus at Xp21.1-p11.4. Roepman et al. (1996) screened this interval for microdeletions using a novel technique they called YAC representation hybridization (YRH). Application of this technique led to the generation of a defined amplifiable subset of restriction fragments representing the insert of a YAC spanning the region of interest. The mixture of PCR products was used to study Southern blots of restriction-digested genomic DNA. In 1 out of 30 patients with X-linked retinitis pigmentosa, they detected a 6.4-kb microdeletion. A cosmid spanning this microdeletion was used to screen cDNA libraries. Roepman et al. (1996) then isolated additional cosmids that flanked the microdeletion region. Shotgun cosmid sequencing enabled them to sequence 32,895 bp of DNA. Computer-assisted analysis of this sequence predicted numerous additional exons which were confirmed by cDNA cloning. Sequence comparisons revealed that the deduced product of the gene showed strong similarity with RCC1, the guanine nucleotide exchange factor of the Ras-like GTPase Ran that is involved in nuclear protein import. Roepman et al. (1996) detected mutations in RP patients and not in controls. Mutation screening was carried out in 28 patients by means of SSCP analysis. They designed intron primers for PCR amplification of 10 exons and detected 5 bandshifts in patients. The corresponding PCR fragments were sequenced and 3 different nucleotide exchanges, and one 4-bp deletion were identified. None of these changes were detected in 84 male controls. The 6 most 3-prime exons showed no mutations but did reveal several polymorphisms. The 3-prime end of the gene is, however, disrupted by the 6.4-kb deletion which is present in a patient with X-linked retinitis pigmentosa. Roepman et al. (1996) noted that the 5-prime end of the gene and the promoter region have not yet been cloned.

In order to characterize the RPGR mutations in a systematic way, Fujita et al. (1997) identified 11 RP3 families by haplotype analysis. Sequence analysis of the PCR-amplified genomic DNA from patients representing these RP3 families showed no causative mutation in RPGR exons 2 to 19, spanning more than 98% of the coding region. In patients from 2 families, however, they identified transition mutations in the intron region near splice sites (IVS10+3; 312610.0005 and IVS13-8). RNA analysis showed that both splice site mutations resulted in the generation of aberrant RPGR transcripts. The results supported the hypothesis that mutations in the RPGR gene are not a common defect in the RP3 subtype of X-linked RP and that the majority of

causative mutations may reside either in as yet unidentified RPGR exons or in another nearby gene at Xp21.1.

Buraczynska et al. (1997) stated that the RPGR gene is mutated in 10 to 15% of European X-linked RP patients and that RP3 is the most frequent genetic subtype of X-linked retinitis pigmentosa. They examined the RPGR gene in a cohort of 80 affected males from apparently unrelated X-linked RP families by direct sequencing of the PCR-amplified products from genomic DNA. Fifteen different putative disease-causing mutations were identified in 17 of the 80 families: 4 nonsense mutations, 1 missense mutation, 6 microdeletions, and 4 intronic-sequence substitutions resulting in splice defects. In their Figure 2, they mapped the location of 12 mutations reported by Meindl et al. (1996) and Roepman et al. (1996) and the 15 different mutations identified in this study. Most of the mutations were detected in the conserved N-terminal region of the RPGR protein, containing tandem repeats homologous to those present in the RCC1 protein. In agreement with previous studies, they were able to demonstrate RPGR mutations in only about 20% of the examined X-linked RP patients. On the other hand, the RP3 subtype consistently accounts for 60 to 90% of families localized by linkage and haplotype genotyping. Buraczynska et al. (1997) raised the possibility that the RPGR gene contains as yet unidentified mutational hotspots in sequences that have not been screened, such as the promoter region or intronic sequences and exon 1. The authors could not rule out the alternative possibility of another gene located in proximity to RPGR at Xp21.1 that also causes RP when mutated.

Souied et al. (1997) described 9 families that showed an X-linked pattern of inheritance with a total of 28 affected males and 34 affected females. The females in these families met criteria for the diagnosis of retinitis pigmentosa. The males had a delayed onset of disease, with central vision being preserved until 40 to 45 years of age. Linkage to the RP3 locus was demonstrated, but SSCP and sequence analysis of the RPGR gene demonstrated no mutations. Souied et al. (1997) suggested that these families demonstrated an X-linked dominant form of RP and that the negative mutation results may be explained either by allelic heterogeneity at the RP3 locus or involvement of a distinct locus mapping close to RP3.

Kirschner et al. (1999) studied the expression of the RPGR gene by Northern blot hybridization, cDNA library screening, and RT-PCR in various organs of mouse and human and identified at least 12 alternatively spliced isoforms. Some of the transcripts are tissue-specific and contain novel exons, which elongate or truncate the previously reported open reading frame of the mouse and human RPGR gene. Kirschner et al. (1999) identified a new exon, designated exon 15A by them, which is expressed exclusively in human retina and mouse eye and contains a premature stop codon. The deduced polypeptide lacked 169 amino acids from the C terminus of the ubiquitously expressed variant, including an isoprenylation site. This exon was deleted in a family with X-linked RP. Kirschner et al. (1999) concluded that their results indicate tissue-dependent regulation of alternative splicing of the RPGR gene and that the presence of the retina-specific transcript may explain why phenotypic aberrations in RP3 are confined to the eye.

The RPGR gene has been shown to be mutated in 10 to 20% of patients with X-linked retinitis pigmentosa. Miano et al. (1999) found a total of 29 different RPGR mutations identified in northern European and United States patients. They performed mutation analysis of the RPGR gene in a cohort of 49 southern European males with XLRP. By multiplex SSCA and direct sequencing of all 19 RPGR exons, 7 different mutations, all novel, were identified in 8 of the 49 families; these included 3 splice site mutations, 2 microdeletions, and 2 missense mutations. RNA analysis showed that the 3 splice site defects resulted in the generation of aberrant RPGR transcripts. Six of these mutations were detected in the conserved N-terminal region of RPGR

protein, containing tandem repeats homologous to repeats within the RCC1 protein (179710). Strikingly, none of the RPGR mutations reported in other populations were identified in this series.

WILD-TYPE

The term "wild-type" is used in its usual sense – i.e. the phenotype that is characteristic of most of the members of a species occurring naturally and contrasting with the phenotype of a mutant (e.g. see Oxford Dictionary of Biochemistry and Molecular Biology, Oxford University Press, 1997).

DISEASE CAUSING MUTATION(S)

The disease causing mutation(s) of the present invention are mutations that are capable of leading to a disease state.

Hence, the disease causing mutation(s) of the present invention are in contrast to polymorphisms. This is because the disease causing mutation(s) are typically present in a small population and are lethal in the sense that their presence will lead to a disease state, whereas in contrast polymorphisms typically occur in larger population percentages and do not necessarily lead to a disease state.

Each of the disease causing mutation(s) of the present invention may be located in a region of a *RPGR* gene. Such a region is termed a mutation hot spot region.

MULTIPLE

The term "multiple" refers to two or more genetically determined alternative sequences or alleles in a population.

ALLELE

The term "allele" refers to a variant form of a gene occuring at a same locus or to different sequence variants found at given markers.

MARKER

The term "marker" refers to a specific site in a gene which exhibits sequence variations between individuals.

SEQUENCE VARIATIONS

The term "sequence variations" includes but is not limited to single or multiple base changes including insertions, deletions or substitutions or a variable number of sequence repeats. As used herein, the terms "sequence variant" and "allele" are used interchangeably with the term "disease causing mutation(s)".

TYPES OF DISEASE CAUSING MUTATIONS

The disease causing mutation(s) may include restriction fragment length mutations, variable number of tandem repeats, single nucleotide mutations, hypervariable regions, minisatellites, dinucleotide repeats, trinucleotide repeats, tetranucleotide repeats, simple sequence repeats, and insertion elements. A disease causing mutation(s) may be as small as one base pair. A one base pair change may occur in a codon.

As used herein, the term "codon" means a sequence of three adjacent nucleotides (a trinucleotide sequence) that may designate an amino acid or a start/stop site for translation.

The disease causing mutation(s) can introduce a number of dirrerent effects – such as the insertion of different amino acid(s) into the expressed protein, the substitution of different amino acid(s) into the expressed protein, the deletion of amino acid(s) from the expressed protein, or the introduction of early stop signals.

Preferred examples of disease causing mutations are presented in the attached sequence listings, in particular see SEQ ID No. 2 and SEQ ID No. 3 and their associated commentary.

RISK ASSOCIATIONS

As used herein, the term "risk association" means that the presence of the disease causing mutation(s) means that the individual was in a very high risk category for that disease state.

Hence, the present invention provides for a method of diagnosing a disease or a predisposition to said disease by genotyping a *RPGR* gene. By genotyping the *RPGR* gene, the methods of the present invention enable either direct diagnosis of a disease or a diagnosis of a predisposition to certain disease conditions.

PHENOTYPE

As used herein, the term "phenotype" means any detectable trait that is the result of one or more genes. A mutation may contribute to the phenotype of an individual in different ways. Some mutations may occur within a protein coding sequence (such as an exon) and contribute to phenotype by affecting protein structure. Other mutations may occur in non coding regions (such as a promoter region or an intron) but may exert phenotypic effects indirectly via influence on replication, transcription, and translation. A single disease causing mutation(s) may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by disease causing mutation(s) in different genes. Further, some disease causing mutation(s) predispose an individual to a distinct mutation that is causally related to a certain phenotype or phenotypic trait.

PHENOTYPIC TRAITS

The disease causing mutation(s) may contribute to the phenotype of an individual in different ways. The disease causing mutation(s) occur within a protein coding sequence and contribute to phenotype by affecting protein structure.

Other disease causing mutation(s) may occur in non coding regions (such as a promoter region or an intron) but may exert phenotypic effects indirectly via influence on replication, transcription, and translation.

A single disease causing mutation(s) may affect more than one phenotypic trait. Likewise, a single phenotypic trait may be affected by disease causing mutation(s) in different genes.

Further, some disease causing mutation(s) predispose an individual to a distinct mutation that is causally related to a certain phenotype.

CORRELATION OF MUTATIONS WITH PHENOTYPIC TRAITS

As an example of the present invention in use (described in more detail below) disease causing mutation(s) in the human *RPGR* gene were identified and their association with risk traits were assessed.

OTHER RISK FACTORS

Optionally, the assessment of an individual's risk factor is calculated by reference also to other known genetic or physiological or dietary or other indications. The invention in this way provides further information on which measurement of an individual's risk of disease or predisposition can be based.

GENOTYPE

As used herein, the term "genotype" means a *RPGR* gene which has been screened for the presence of at least one disease causing mutation(s) at a specific genetic locus. Otherwise expressed, the screened *RPGR* gene could be called a "genotyped *RPGR* gene".

RISK GENOTYPE

As used herein, the term "risk genotype" refers to a *RPGR* gene which comprises at least one disease causing mutation(s) which is associated with at least one disease phenotype or phenotypic trait.

GENOTYPING

As used herein, the term "genotyping" means determining whether a *RPGR* gene includes at least one disease causing mutation(s). The term "genotyping" is synonymous with terms such as "genetic testing", "genetic screening", "determining or identifying an allele", "molecular diagnostics" or any other similiar phrase.

Any method capable of distinguishing nucleotide differences in the appropriate sample DNA sequences may also be used. In fact, a number of known different methods are suitable for use in genotyping (that is, determining the genotype) for a mutant *RPGR* gene of the present invention. These methods include but are not limited to direct sequencing, PCR-RFLP, ARMS-PCR, Taqman™, Molecular beacons, hybridization to oligonucleotides on DNA chips and arrays, single nucleotide primer extension and oligo ligation assays.

GENOTYPE SCREENING

In one embodiment, the present invention provides a method for genotype screening of a nucleic acid comprising a *RPGR* gene from an individual. The methods for genotype

screening of a nucleic acid comprising a *RPGR* gene from an individual may require amplification of a nucleic acids from a target sample from that individual.

GENOTYPING MUTATIONS

A number of different methods are suitable for use in determining the genotype for a mutation. These methods include but are not limited to direct sequencing, PCR-RFLP, ARMS-PCR, Taqman™, Molecular beacons, hybridization to oligonucleotides on DNA chips and arrays, single nucleotide primer extension and oligo ligation assays. Any method capable of distinguishing single nucleotide differences in the appropriate DNA sequences may also be used.

DISEASE STATE

The present application provides *inter alia* a means for detecting a certain disease state. Here, the disease state is typically XLRP.

PREDISPOSITION TO DISEASE

As used herein, the term "predisposition to a disease" means that certain disease causing mutation(s) are shown to be associated with a given disease state. They are thus represented in individuals with disease as compared with healthy individuals and indicate that these individuals are at a very high risk for developing disease or may develop a more severe form or particular subset of the disease type.

DIAGNOSIS OF DISEASE

The methods of diagnosis of predisposition to the disease state involve determining whether an individual possesses the published wild-type sequence or the disease causing mutation(s) at one or more of the disease causing mutation(s). In this respect, the genotype of the individual is compared with the phenotype of the individual. As

used herein, the term "phenotype" means any detectable trait of an individual that is the result of one or more genes.

MUTANT *RPGR* GENE

The present invention is therefore concerned with hitherto unrecognised disease causing mutation(s) in a wild-type RPGR gene. For convenience, these mutant sequences are present in sequences that are collectively refererred to as being a "mutant *RPGR* gene".

For convenince, the term "mutant *RPGR* gene" as used herein includes references to one or more of said sequences presented herein, or a variant, homologue or derivative of any one or more thereof.

The term "mutant *RPGR* gene" also includes references to fragments one or more of said sequences presented herein, or a variant, homologue or derivative of any one or more thereof. Hence, the term mutant *RPGR* gene includes references to any one of the sequences presented as SEQ ID No.s 1, 2 or 3.

Preferably said variants, homologues, derivatives or fragments comprise one or more of the disease causing mutation(s) mentioned herein.

Likewise, the present invention is concerned with hitherto unrecognised mutant RPGR proteins associated with the disease causing mutation(s) in a wild-type *RPGR* gene. For convenience, these mutant sequences are collectively refererred to as being a "mutant RPGR protein". Here, the term "mutant protein" includes references to one or more of said sequences presented herein, or a variant, homologue or derivative of any one or more thereof.

The term "mutant RPGR protein" also includes references to fragments one or more of said sequences presented herein, or a variant, homologue or derivative of any one or more thereof.

Preferably said variants, homologues, derivatives or fragments comprise one or more amino acids associated with one or more of the the disease causing mutation(s) mentioned herein.

Wherever appropriate, the term "mutant *RPGR* gene" may be used interchangeably with the gene coding for same - otherwise expressed as being a nucleotide sequence of interest (NOI) – and/or any biologically active fragment(s) thereof and/or the expression product thereof – otherwise expressed as EP and/or any biologically active fragment(s) thereof.

Likewise, wherever appropriate, the term "*RPGR* gene" may be used interchangeably with the gene coding for same - otherwise expressed as being a nucleotide sequence of interest (NOI) – and/or any biologically active fragment(s) thereof and/or the expression product thereof – otherwise expressed as EP and/or any biologically active fragment(s) thereof.

The term "NOI" includes DNA, RNA and single and double stranded sequences. It also refers to sequences which are prepared by synthetic means.

For some applications, the NOI is in an isolated and/or purified form.

For some applications, the EP is in an isolated and/or purified form.

ISOLATED MUTANT *RPGR* GENE

The isolated mutant *RPGR* gene of the present invention may be introduced into a vector and expressed under *in vitro*, and/or *in vivo* and/or *ex vivo* conditions. In this way, the expression product may be used in applications which include but are not limited to gene therapy, identification of potential pharmaceutical targets in high throughput screening (HTS) assays and forensic analysis.

For some aspects of the present invention, preferably the isolated mutant *RPGR* gene of the present invention is introduced into a vector and expressed under *in vitro*, and/or *in vivo* and/or *ex vivo* conditions.

The nucleotide sequences of the invention may be in a substantially isolated form. It will be understood that the nucleotide sequence may be mixed with carriers or diluents which will not interfere with the intended purpose of the protein and still be regarded as substantially isolated. A nucleotide sequence of the invention may also be in a substantially purified form, in which case it will generally comprise the nucleotide sequence in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the preparation is a nucleotide sequence of the present invention.

EP ISOLATION

The expression product (EP) of the nucleotide sequences of the present invention may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, i.e., 80, 95 or 99% free of cell component contaminants, as described in Jacoby, Methods in Enzymology Volume 104, Academic Press, New York (1984); Scopes, Protein Purification, Principles and Practice, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), Guide to Protein Purification, Methods in Enzymology, Vol. 182 (1990). If the EP is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the EP can be isolated from a lysate of the host cells.

Proteins of the invention may be in a substantially isolated form. It will be understood that the protein may be mixed with carriers or diluents which will not interfere with the intended purpose of the protein and still be regarded as substantially isolated. A protein of the invention may also be in a substantially purified form, in which case it will generally comprise the protein in a preparation in which more than 90%, e.g. 95%, 98% or 99% of the protein in the preparation is a protein of the invention.

REFERENCE SEQUENCE

As used herein the term "reference sequence" means an amino acid sequence or a nucleotide sequence – typically a nucleotide sequence - representing one or more individuals homozygous for each of the alleles being tested, such as in a diagnostic assay. These reference sequences may be called control sequences or reference samples or control samples.

By way of example, reference DNA sequences may include but are not limited to: (i) a genomic DNA from homozygous individuals; (ii) a PCR product containing a relevant mutation amplified from homozygous individuals; or (iii) a DNA sequence containing a relevant mutation that has been cloned into a plasmid or other suitable vector.

The reference sample may also be an alleleic ladder comprising a plurality of alleles from known set of alleles. There may be a plurality of reference samples, each containing different alleles or sets of alleles. Other reference samples typically include diagrammatic representations, written representations, templates or any other means suitable for identifying the presence of one or more the disease causing mutation(s) in a PCR product or other fragment of nucleic acid.

TARGET SAMPLE

The target sample of the present invention may be any target nucleic acid comprising a *RPGR* gene, and in particular a mutant *RPGR* gene. The target may be for diagnostic purposes and/or analytical purposes. The target sequence is typically obtained from an individual being analyzed.

For assay using these nucleic acids, virtually any biological sample is suitable. For example, convenient target samples include but are not limited to whole blood, leukocytes, semen, saliva, tears,urine, fecal material, sweat, buccal, skin and hair. For

an assay of cDNA or mRNA, the target sample is typically obtained from a cell or organ in which the target nucleic acid is expressed.

In some circumstances, the target sample of the present invention may be any target aminio acid acid comprising the expression product of the *RPGR* gene, or part thereof, from an individual being analyzed.

For the screeing assays of the present invention the target sample may be the mutant *RPGR* gene. However, such assays may also utilise the wild-type *RPGR* gene.

In addition, or in the alternative, for the screeing assays of the present invention the target sample may be the expression product of the mutant *RPGR* gene. However, such assays may also utilise the expression product of the wild-type *RPGR* gene.

NUCLEOTIDE SEQUENCE

The present invention provides novel nucleotide sequences associated with certain the disease causing mutation(s) of the *RPGR* gene. The present invention also relates to novel fragments of those nucleotide sequences. Here, the term "nucleotide sequence" includes sequences having at least more than 5, 10 or 20 bases.

For convenience, the nucleotide sequences of the present invention (or fragments thereof) are sometimes referred to as being mutant *RPGR* gene.

The term mutant *RPGR* gene also encompasses variants, homologues or derivatives of the sequences presented herein.

In particular, the term mutant *RPGR* gene encompasses variants, homologues or derivatives of the sequences presented as SEQ ID No. 1 or 2.

Where the polynucleotide of the invention is double-stranded, both strands of the duplex, either individually or in combination, are encompassed by the present invention. Where

the polynucleotide is single-stranded, it is to be understood that the complementary sequence of that polynucleotide is also included within the scope of the present invention.

Polynucleotides of the invention may be used to produce a primer, e.g. a PCR primer, a primer for an alternative amplification reaction, a probe e.g. labelled with a revealing label by conventional means using radioactive or non-radioactive labels, or the polynucleotides may be cloned into vectors. Such primers, probes and other fragments will be at least 15, preferably at least 20, for example at least 25, 30 or 40 nucleotides in length, and are also encompassed by the term polynucleotides of the invention as used herein.

Polynucleotides such as a DNA polynucleotides and probes according to the invention may be produced recombinantly, synthetically, or by any means available to those of skill in the art. They may also be cloned by standard techniques.

In general, primers will be produced by synthetic means, involving a step wise manufacture of the desired nucleic acid sequence one nucleotide at a time. Techniques for accomplishing this using automated techniques are readily available in the art.

Longer polynucleotides will generally be produced using recombinant means, for example using a PCR (polymerase chain reaction) cloning techniques. This will involve making a pair of primers (e.g. of about 15 to 30 nucleotides) flanking a region of the lipid targeting sequence which it is desired to clone, bringing the primers into contact with mRNA or cDNA obtained from an animal or human cell, performing a polymerase chain reaction under conditions which bring about amplification of the desired region, isolating the amplified fragment (e.g. by purifying the reaction mixture on an agarose gel) and recovering the amplified DNA. The primers may be designed to contain suitable restriction enzyme recognition sites so that the amplified DNA can be cloned into a suitable cloning vector.

AMINO ACID SEQUENCE

The present invention provides novel amino acid sequences associated with certain the disease causing mutation(s) forms of the *RPGR* gene. The present invention also relates to novel fragments of those amino acid sequences.

The amino acid sequences are sometimes referred to as proteins. Here, the term "protein" includes polypeptides having at least more than 5, 10 or 20 amino acids.

For convenience, the amino acid sequences of the present invention (or fragments thereof) are sometimes referred to as being mutant RPGR protein.

The term mutant RPGR protein also encompasses variants, homologues or derivatives of the sequences presented herein.

In particular, the term mutant RPGR protein encompasses variants, homologues or derivatives of the sequences presented as SEQ ID No. 1 or 2, or fragments thereof.

VARIANTS/HOMOLOGUES/DERIVATIVES

It will be understood that sequences of the invention or for use in the invention are not limited to the particular sequences or fragments thereof or sequences obtained from the particular sources mentioned herein but also include homologous sequences obtained from any source, for example related proteins, cellular homologues and synthetic peptides, as well as variants or derivatives thereof.

Thus, the present invention covers variants, homologues or derivatives of the protein sequences of the present invention, as well as variants, homologues or derivatives of the nucleotide sequence coding for the protein sequences of the present invention.

Thus, in addition to the specific amino acid sequences and nucleotide sequences mentioned herein, the present invention also encompasses the use of variants,

homologue and derivatives thereof. Here, the term "homology" can be equated with "identity".

In the present context, an homologous sequence is taken to include a sequence which may be at least 75, 85 or 90% identical, preferably at least 95 or 98% identical. In particular, homology should typically be considered with respect to those regions of the sequence known to be essential for an activity. Although homology can also be considered in terms of similarity (i.e. amino acid residues having similar chemical properties/functions), in the context of the present invention it is preferred to express homology in terms of sequence identity.

Homology comparisons can be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs can calculate % homology between two or more sequences.

% homology may be calculated over contiguous sequences, i.e. one sequence is aligned with the other sequence and each amino acid in one sequence is directly compared with the corresponding amino acid in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues.

Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion will cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without penalising unduly the overall homology score. This is achieved by inserting "gaps" in the sequence alignment to try to maximise local homology.

However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence

alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - will achieve a higher score than one with many gaps. "Affine gap costs" are typically used that charge a relatively high cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties will of course produce optimised alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example when using the GCG Wisconsin Bestfit package (see below) the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension.

Calculation of maximum % homology therefore firstly requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A.; Devereux et al., 1984, Nucleic Acids Research 12:387). Examples of other software than can perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 ibid – Chapter 18), FASTA (Atschul et al., 1990, J. Mol. Biol., 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999 ibid, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program. A new tool, called BLAST 2 Sequences is also available for comparing protein and nucleotide sequence (see FEMS Microbiol Lett 1999 174(2): 247-50; FEMS Microbiol Lett 1999 177(1): 187-8 and tatiana@ncbi.nlm.nih.gov).

Although the final % homology can be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pairwise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table if supplied (see user manual for

further details). It is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62.

Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result.

The sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent substance. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the secondary binding activity of the substance is retained. For example, negatively charged amino acids include aspartic acid and glutamic acid; positively charged amino acids include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values include leucine, isoleucine, valine, glycine, alanine, asparagine, glutamine, serine, threonine, phenylalanine, and tyrosine.

Conservative substitutions may be made, for example according to the Table below. Amino acids in the same block in the second column and preferably in the same line in the third column may be substituted for each other:

ALIPHATIC	Non-polar	G A P
		I L V
	Polar - uncharged	C S T M
		N Q
	Polar - charged	D E
		K R
AROMATIC		H F W Y

The present invention also encompasses homologous substitution (substitution and replacement are both used herein to mean the interchange of an existing amino acid residue, with an alternative residue) may occur i.e. like-for-like substitution such as basic for basic, acidic for acidic, polar for polar etc. Non-homologous substitution may also occur i.e. from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine, a more detailed list of which appears below.

Replacements may also be made by unnatural amino acids include; alpha* and alpha-disubstituted* amino acids, N-alkyl amino acids*, lactic acid*, halide derivatives of natural amino acids such as trifluorotyrosine*, p-Cl-phenylalanine*, p-Br-phenylalanine*, p-I-phenylalanine*, L-allyl-glycine*, ß-alanine*, L-α-amino butyric acid*, L-γ-amino butyric acid*, L-α-amino isobutyric acid*, L-ε-amino caproic acid#, 7-amino heptanoic acid*, L-methionine sulfone#*, L-norleucine*, L-norvaline*, p-nitro-L-phenylalanine*, L-hydroxyproline#, L-thioproline*, methyl derivatives of phenylalanine (Phe) such as 4-methyl-Phe*, pentamethyl-Phe*, L-Phe (4-amino)#, L-Tyr (methyl)*, L-Phe (4-isopropyl)*, L-Tic (1,2,3,4-tetrahydroisoquinoline-3-carboxyl acid)*, L-diaminopropionic acid # and L-Phe (4-benzyl)*. The notation * has been utilised for the purpose of the discussion above (relating to homologous or non-homologous substitution), to indicate the hydrophobic nature of the derivative whereas # has been utilised to indicate the hydrophilic nature of the derivative, #* indicates amphipathic characteristics.

Variant amino acid sequences may include suitable spacer groups that may be inserted between any two amino acid residues of the sequence including alkyl groups such as methyl, ethyl or propyl groups in addition to amino acid spacers such as glycine or β-alanine residues. A further form of variation, involves the presence of one or more amino acid residues in peptoid form, will be well understood by those skilled in the art. For the avoidance of doubt, "the peptoid form" is used to refer to variant amino

acid residues wherein the α-carbon substituent group is on the residue's nitrogen atom rather than the α-carbon. Processes for preparing peptides in the peptoid form are known in the art, for example Simon RJ et al., PNAS (1992) 89(20), 9367-9371 and Horwell DC, Trends Biotechnol. (1995) 13(4), 132-134.

Polynucleotides which are not 100% homologous to the sequences of the present invention but fall within the scope of the invention can be obtained in a number of ways. Other variants of the sequences described herein may be obtained for example by probing DNA libraries made from a range of individuals, for example individuals from different populations. In addition, other viral/bacterial, or cellular homologues particularly cellular homologues found in mammalian cells (e.g. rat, mouse, bovine and primate cells), may be obtained and such homologues and fragments thereof in general will be capable of selectively hybridising to the sequences shown in the sequence listing herein. Such sequences may be obtained by probing cDNA libraries made from or genomic DNA libraries from other animal species, and probing such libraries with probes comprising all or part of the sequences present herein (especially those that comprise the the disease causing mutation(s) regions) under conditions of medium to high stringency. Similar considerations apply to obtaining species homologues and allelic variants of the polypeptide sequences of the invention.

Variants and strain/species homologues may also be obtained using degenerate PCR which will use primers designed to target sequences within the variants and homologues encoding conserved amino acid sequences within the sequences of the present invention. Conserved sequences can be predicted, for example, by aligning the amino acid sequences from several variants/homologues. Sequence alignments can be performed using computer software known in the art. For example the GCG Wisconsin PileUp program is widely used.

The primers used in degenerate PCR will contain one or more degenerate positions and will be used at stringency conditions lower than those used for cloning sequences with single sequence primers against known sequences.

Alternatively, such polynucleotides may be obtained by site directed mutagenesis of characterised sequences. This may be useful where for example silent codon changes are required to sequences to optimise codon preferences for a particular host cell in which the polynucleotide sequences are being expressed. Other sequence changes may be desired in order to introduce restriction enzyme recognition sites, or to alter the property or function of the polypeptides encoded by the polynucleotides.

BIOLOGICALLY ACTIVE FRAGMENTS

In addition to substantially full-length EPs (such as a polypeptide) expressed by NOIs of the present invention, the EPs of the present invention may include biologically active fragments, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confer a biological function on the EP, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

FUSION PROTEINS

Proteins of the invention are typically made by recombinant means, for example as described below. However they may also be made by synthetic means using techniques well known to skilled persons such as solid phase synthesis. Proteins of the invention may also be produced as fusion proteins, for example to aid in extraction and purification. Examples of fusion protein partners include glutathione-S-transferase (GST), 6xHis, GAL4 (DNA binding and/or transcriptional activation domains) and β-galactosidase. It may also be convenient to include a proteolytic cleavage site between the fusion protein partner and the protein sequence of interest to allow removal of fusion protein sequences. Preferably the fusion protein will not hinder the function of the EP.

PROBES/PRIMERS

The present invention also provides a series of useful probes – otherwise known as primers.

As used herein, the term "primer" refers to a single-stranded oligonucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. A primer need not reflect the exact sequence of the template but must be sufficiently complementary to hybridize with a template.

The term "primer site" refers to the area of the target DNA to which a primer hybridizes.

The term "primer pair" means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3' downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

The primers of the present invention may be DNA or RNA, and single-or double-stranded. Alternatively, the primers may be naturally occurring or synthetic, but are typically prepared by synthetic means.

ALLELE SPECIFIC PROBES/PRIMERS

An allele-specific primer hybridizes to a site on target DNA overlapping a disease causing mutation(s) and only primes amplification of an allelic form to which the

primer exhibits at least substantially perfect complementarity (See Gibbs, Nucleic Acid Res. 17, 2427-2448 (1989)). This primer may be used in conjunction with a second primer which hybridizes at a distal site. Amplification proceeds from the two primers leading to a detectable product signifying the particular allelic form is present. A control may be performed with a second pair of primers, one of which shows a single base mismatch at the mutant site and the other of which exhibits perfect complementarily to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the disease causing mutation(s) because this position is most destabilizing to elongation from the primer (see, for example WO 93/22456). Hybridisation probes capable of specific hybridisation to detect a single base mismatch may be designed according to methods known in the art and described in Maniatas *et al* Molecular Cloning: A Laboratory Manual, 2nd Ed (1989) Cold Spring Harbour.

Hence, allele-specific probes can be designed that hybridize to a segment of target DNA from one individual but do not hybridize to the corresponding segment from another individual due to the presence of different forms in the respective segments from the two individuals.

As used herein, the term "probe" refers to an oligonucleotide (ie a sequence of nucleotides), whether occuring naturally as in a purified restriction digest or produced synthetically, which is capable of hybridising to another oligonucleotide sequence of interest. Probes are useful in the detection, identification and isolation of particular gene sequences. The hybridization probes of the present invention are typically oligonucleotides capable of binding in a base-specific manner to a complementary strand of nucleic acid.

The probes of the present invention may be labelled with any "reporter molecule" so that it is detectable in any detection system, including but not limited to enzyme (for example, ELISA, as well as enzyme based histochemical assays), fluorescent, radioactive and luminescent systems. The target sequence of interest (that is, the

sequence to be detected) may also be labelled with a reporter molecule. The present invention is not limited to any particular detection system or label.

The hybridization conditions chosen for the probes of the present invention are sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the alleles. The typical hybridisation conditions are stringent conditions as set out above for the allele specific primers of the present invention so that a one base pair mismatch may be determined.

PCR ALLELE SPECIFIC PRIMERS

Preferably the screening is carried out using PCR allele specific primers designed to amplify portions of the *RPGR* gene that include one or more of the disease causing mutation(s).

Examples of such PCR primers are based on the sequences presented herein, in particular those based on sequence mutations presented in Table 1 or the sequences presented in or as SEQ ID No. 2 or SEQ ID No. 3.

HYBRIDISATION

As used herein, the term "hybridisation" refers to the pairing of complementary nucleic acids. Hybridisation and the strength of hybridisation (ie the strength of association between the nucleic acids) is impacted by such factors as the degree of complementarity between nucleic acids, stringency of conditions involved, the melting temperature (Tm) of the formed hybrid and the G:C ratio within the nucleic acids.

As used herein, the term "stringency" is used in reference to the conditions of temperature, ionic strength and the presence of other compounds such as organic solvents under which the nucleic acid hybridisations are conducted.

Hybridizations are typically performed under stringent conditions, for example, at a salt concentration of no more than 1M and a temperature of at least 25°C. For example, conditions of 5X SSPE (750 mM NaCl, 50 mM NaPhosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific primer hybridizations.

AMPLIFICATION

As used herein, the term "amplification means nucleic acid replication involving template specificity. The template specificity relates to a "target sample" or "target sequence" specificity. The target sequences are "targets" in the sense that they are sought to be sorted out from other nucleic acids. Consequently, amplification techniques have been designed primarily for sorting this out. Examples of amplification methods include but are not limited to polymerase chain reaction (PCR), polymerase chain reaction of specific alleles (PASA), ligase chain reaction (LCR), transcription amplification, self-sustained sequence replication and nucleic acid based sequence amplification (NASBA).

TAQMAN™

Suitable means for determining genotype may be based on the Taqman™ technique. The Taqman™ technique is disclosed in the following US patents 4,683,202; 4,683,195 and 4,965,188. The use of uracil N-glycosylase which is included in Taqman™ allelic discrimination assays is disclosed in US patent 5,035,996.

PCR

PCR techniques are well known in the art (see for example, EP-A-0200362 and EP-A-0201184 and US patent Nos 4 683 195 and 4 683 202). The process for amplifying the target sequence consists of introducing a large excess of two oligonucleotide primers to the DNA mixture containing the desired target sequence, followed by a precise sequence of thermal cycling in the presence of a DNA polymerase. With PCR, it is possible to

amplify a single copy of a specific target sequence in, for example, genomic DNA to a level detectable by several different methodologies (such as hybridisation with a labelled probe, incorporation of biotinylated primers followed by avidin-enzyme conjugate detection and incorporation of ^{32}P labelled deoxynucleotide triphosphates, such as dCTP or dATP, into the amplified sequence). Alternatively, it is possible to amplify different disease causing mutation(s) (markers) with primers that are differentially labelled and thus can each be detected. One means of analysing multiple markers involves labelling each marker with a different fluorescent probe. The PCR products are then analysed on a fluorescence based automated sequencer. In addition to genomic DNA, any oligonucleotide sequence may be amplified with the appropriate set of primer molecules. In particular, the amplified segments created by the PCR process itself are, themselves, efficient templates for subsequent PCR amplifications. By way of example, PCR can also be used to identify primers for amplifying suitable sections of a *RPGR* gene in or from a human.

DIAGNOSTIC KITS

The present invention also provides for a kit for diagnosis of or predisposition to disease, said kit typically comprising: (a) means for determining the genotype of a *RPGR* gene in a human; and (b) reference means for identifying the presence of a disease causing mutation(s).

Typically, the kit of the present inventon contains all of the necessary components to determine the presence/absence of a disease causing mutation(s) of the present invention in an individual. These components include, but are not limited to, PCR primers, PCR enzymes, restriction enzymes, a DNA purification means, a DNA sampling means and any other component useful for determining a mutational difference between the wildtype *RPGR* gene and an allelic *RPGR* variant of the present invention. By way of example, the kits may comprise at least one allele-specific oligonucleotide primer and/or allele-specific oligonucleotide probe. Alternatively, the kits contain one or more pairs of allele-specific oligonucleotides capable of hybridizing to different forms of a mutation(s) – such as that found in the

RPGR gene. In some kits, the allele-specific oligonucleotides may be immobilized to a substrate. By way of example, the same substrate can comprise allele-specific oligonucleotide probes for detecting at least the disease causing mutation(s). Optional additional components of the kit may include, for example, means used to label (for example, an avidin enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse transcription, PCR, or hybridization reactions. The control/reference sample may comprise a wild type *RPGR* gene or may contain an allele known to be associated with an age-related disease. Alternatively, the reference/control sample may comprise actual PCR products produced by amplification of relevant disease related alleles or may contain genomic or cloned DNA from an individual with a known set of particular disease related alleles. Usually, the kit also contains instructions for carrying out the methods. The kit may also contain a modulator capable of overcoming the disease causing mutation(s). The kits may be used for detection and measurement of the disease causing mutation(s) in biological fluids and tissues, and for localization of a mutation in tissues. The kits may also be used in simultaneously or sequentially with an agent (such as a modulator) as defined herein.

DETECTION OF DISEASE CAUSING MUTATIONS IN AMPLIFIED TARGET SEQUENCES

The amplified nucleic acid sequences may be detected using procedures including but not limited to allele-specific probes, tiling arrays, direct sequencing, denaturing gradient gel electrophoresis and single-strand conformation polymorphism (SCCP) analysis. However, in the present case it would be more appropriate to call it single-strand conformation disease causing mutation(s) (SCCDCM) analysis.

TILING ARRAYS

The disease causing mutation(s) of the present invention may also be identified by hybridization to nucleic acid arrays, some example of which are described in WO 95/11995. The term "tiling" generally means the synthesis of a defined set of

oligonucleotide probes that is made up of a sequence complementary to the sequence to be analysed (the "target sequence"), as well as preselected variations of that sequence. The variations usually include substitution at one or more base positions with one or more nucleotides.

DIRECT SEQUENCING

The direct analysis of the sequence of the disease causing mutation(s) of the present invention may be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook *et al.*, Molecular Cloning, A Laboratory Manual (2nd Ed., CSHP, New York 1989) or using, for example, Standard ABI sequencing technology using Big Dye Terminator cycle sequencing chemistry analyzed on an ABI Prism 377 DNA sequencer.

DENATURING GRADIENT GEL ELECTROPHORESIS

Amplification products of the present invention, which are generated using PCR, may also be analyzed by the use of denaturing gradient gel electrophoresis. Different alleles may be identified based on the different sequence-dependent melting properties and electrophoretic migration of DNA in solution. Erlich, ed., PCR Technology, Principles and Applications for DNA Amplification, (W.H. Freeman and Co, New York, 1992), Chapter 7.

SINGLE-STRAND CONFORMATION POLYMORPHISM (SCCP) ANALYSIS

The amplified nucleic acid sequences may be detected using single-strand conformation polymorphism (SCCP) analysis; but as indicated above in the present case it would be more appropriate to call it single-strand conformation disease causing mutation(s) (SCCDCM) analysis.

Alleles of target sequences of the present invention may also be differentiated using SCCP analysis (however, in the present application one is identifying disease causing

mutation(s) and not polymorphisms – nevertheless, this particular technology is still applicable for this application), which identifies base differences by alteration in electrophoretic migration of single stranded PCR products, as described in Orita *et al.*, Proc. Nat. Acad. Sci. 86, 2766-2770(1989). Amplified PCR products can be generated as described above, and heated or otherwise denatured, to form single stranded amplification products. Single-stranded nucleic acids may refold or form secondary structures which are partially dependent on the base sequence. The different electrophoretic mobilities of single-stranded amplification products may be related to base-sequence difference between alleles of target sequences.

IDENTIFYING DIFFERENCES BETWEEN TEST AND CONTROL SEQUENCES

Typical detection procedures for amplified nucleic acid sequences may be used to identify difference of one or more points of varation between a reference and test nucleic acid sequence or to compare different disease causing mutation(s) forms of the RPGR gene from two or more individuals.

VECTORS

As it is well known in the art, a vector is a biological tool that allows or faciliates the transfer of an entity from one environment to another. Examples of vectors used in recombinant DNA techniques include but are not limited to plasmids, chromosomes, artificial chromosomes or viruses.

The term "vector" includes expression vectors and/or transformation vectors.

The term "expression vector" means a construct capable of *in vivo* or *in vitro/ex vivo* expression.

The term "transformation vector" means a construct capable of being transferred from one species to another.

EXPRESSION VECTOR

Preferably, the nucleotide sequence of interest (NOI) which is inserted into a vector is operably linked to a control sequence that is capable of providing for the expression of the coding sequence by the host cell, i.e. the vector is an expression vector. The expression product (EP) produced by a host recombinant cell may be secreted or may be contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing the NOI can be designed with signal sequences which direct secretion of the NOI coding sequences through a particular prokaryotic or eukaryotic cell membrane.

VECTOR TRANSFER

The vectors comprising nucleotide sequences (NOIs) of the present invention may be introduced into suitable host cells using a variety of techniques known in the art, such as transfection, transformation, electroporation and biolistic transformation.

As used herein, the term "transfection" refers to a process using a non-viral vector to deliver a gene to a target mammalian cell.

Typical transfection methods include electroporation, DNA biolistics, lipid-mediated transfection, compacted DNA-mediated transfection, liposomes, immunoliposomes, lipofectin, cationic agent-mediated, cationic facial amphiphiles (CFAs) (Nature Biotechnology 1996 14; 556), multivalent cations such as spermine, cationic lipids or polylysine, 1, 2,-bis (oleoyloxy)-3-(trimethylammonio) propane (DOTAP)-cholesterol complexes (Wolff and Trubetskoy 1998 Nature Biotechnology 16: 421) and combinations thereof.

HOST CELLS

A wide variety of host cells can be employed for expression of the NOIs of the present invention, both prokaryotic and eukaryotic. Suitable host cells include bacteria

such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, e.g., mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the NOI expression products (EPs) to produce an appropriate mature polypeptide. Processing includes but is not limited to glycosylation, ubiquitination, disulfide bond formation and general post-translational modification.

TRANSGENIC ANIMALS

The invention further provides transgenic nonhuman animals capable of expressing the NOI of the present invention and/or having one or more of the NOIs inactivated and/or removed. Expression of an NOI is usually achieved by operably linking the NOI to a promoter and optionally an enhancer, and microinjecting the construct into a zygote. See Hogan *et al.*, "Manipulating the Mouse Embryo, A Laboratory Manual," Cold Spring Harbor Laboratory. Inactivation of NOIs can be achieved by forming a transgene in which a cloned NOI is inactivated by insertion of a positive selection marker. See Capecchi, Science 244, 1288-1292 (1989). The transgene is then introduced into an embryonic stem cell, where it undergoes homologous recombination with an endogenous variant gene. Mice and other rodents are preferred animals. Such animals provide useful drug screening systems.

REGULATION OF EXPRESSION *IN VITRO/ IN VIVO/EX VIVO*

The present invention also encompasses gene therapy whereby the NOI is regulated *in vitro/in vivo/ex vivo*. For example, expression regulation may be accomplished by administering compounds that bind to NOI or control regions associated with the NOI, or its corresponding RNA transcript to modify the rate of transcription or translation.

CONTROL SEQUENCES

Control sequences that may be operably linked to sequences encoding the NOI include promoters/enhancers and other expression regulation signals. These control sequences may be selected to be compatible with the host cell and/or target cell in which the expression vector is designed to be used. The control sequences may be modified, for example by the addition of further transcriptional regulatory elements to make the level of transcription directed by the control sequences more responsive to transcriptional modulators.

OPERABLY LINKED

The term "operably linked" means that the components described are in a relationship permitting them to function in their intended manner. A regulatory sequence "operably linked" to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under condition compatible with the control sequences.

The NOIs of the present invention can be expressed in an expression vector in which a variant gene is operably linked to a native promoter or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors may also be used. Vectors may also include but are not limited to host-recognized replication systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome.

PROMOTER

As used herein, the term "promoter" refers to a segment of DNA that contains the start signals for RNA polymerase and hence promotes transcription at the start of a structural gene. It also comprises the binding site of transcription factors that regulate gene expression. The promoter DNA segment is typically located in a region 5' to a

structural gene. That is, the promoter DNA segment is typically located it is located in a 5'region.

EXON

As used herein, the term "exon" means any segment of an interrupted gene that is represented in the mature RNA product. By way of example, an exon may be a region within a gene that codes for a polypeptide chain or domain. Typically, a mature protein is composed of several domains coded by different exons within a single gene.

INTRON

As used herein, the term "intron" refers to a segment of an interrupted gene that is not represented in the mature RNA product. Introns are part of the primary nuclear transcript but are spliced out to produce mRNA, which is then transported to the cytoplasm.

5' REGION

As used herein, the term "5' region" means a region which is 5' to a first exon of a structural gene such as a *RPGR* gene. The term "5' region" includes but is not limited to regions such as a 5' non-coding region and putative promoter regions or regions comprising promoter elements.

3' REGION

As used herein, the term "3' region" means a region which is remote from the 5' region.

SCREENS/ASSAYS

The NOIs of the present invention and/or a cell line that expresses the NOIs of the present invention may be used to screen for agents capable of affecting the expression of the sequences and/or the biological activity of the EPs thereof.

As used herein, the term "agent" may include but is not limited to a chemical compound, a mixture of chemical compounds, peptides, organic or inorganic molecules a biological macromolecule, or an extract made from biological materials such as bacteria, fungi, or animal (particularly mammalian) cells or tissues.

In one embodiment, the screens of the present invention may identify agonists and/or antagonists of the expression product of the present invention.

In another embodiment, the NOIs of the present invention may be used in a variety of drug screening techniques. By way of example, the NOI or EP thereof to be employed in such a test may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The abolition of binding specificity/biological activity or the formation of binding complexes between the NOI and/or EP thereof and the agent being tested may be measured.

Another technique for screening provides for high throughput screening (HTS) of agents having suitable binding affinity for the NOI of the present invention and is based upon the method described in detail in WO 84/03564.

It is expected that the assay methods of the present invention will be suitable for both small and large-scale screening of test compounds as well as in quantitative assays.

The invention further provides for a method of identifying a compound to prevent and/or delay and/or reduce and/or treat a disease and/or a predisposition to a disease comprising the steps of: (a) administring a compound to an animal tissue; and (b) determining whether said compound modulates an NOI of the present invention.

In an example of the invention in use a range of compounds are administered to animal cells in tissue culture (*in vitro*). In this way, thousands of potential small molecule modulators of *RPGR* gene expression (such as mutant *RPGR* gene expression) and function can be screened. Modulation of the *RPGR* gene (such as the mutant *RPGR* gene) can either be determined by assessing whether the candidate compound affects levels of expression of a *RPGR* gene (such as the mutant *RPGR* gene), or whether *RPGR* gene (such as the mutant *RPGR* gene) expression product (EP) function is affected. The invention also provides for administration of candidate compounds, that may modulate a *RPGR* gene (such as the mutant *RPGR* gene), to animal tissues *in vivo*, i.e. administration of compounds to live animals and then assessing their effects by routine methods such as histopathological analysis of tissues.

REPORTERS

A wide variety of reporters may be used in the assay methods (as well as screens) of the present invention with preferred reporters providing conveniently detectable signals (eg. by spectroscopy). By way of example, a reporter gene may encode an enzyme which catalyses a reaction which alters light absorption properties.

Examples of reporter molecules include but are not limited to β-galactosidase, invertase, green fluorescent protein, luciferase, chloramphenicol, acetyltransferase, β-glucuronidase, exo-glucanase and glucoamylase. Alternatively, radiolabelled or fluorescent tag-labelled nucleotides can be incorporated into nascent transcripts which are then identified when bound to oligonucleotide probes.

In one preferred embodiment, the production of the reporter molecule is measured by the enzymatic activity of the reporter gene product, such as β-galactosidase.

A variety of protocols for detecting and measuring the expression of the target, such as by using either polyclonal or monoclonal antibodies specific for the protein, are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilising monoclonal antibodies reactive to two non-interfering epitopes is preferred, but a competitive binding assay may be employed. These and other assays are described, among other places, in Hampton R *et al* (1990, Serological Methods, A Laboratory Manual, APS Press, St Paul MN) and Maddox DE *et al* (1983, J Exp Med 15 8:121 1).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic and amino acid assays. Means for producing labelled hybridisation or PCR probes for detecting the target polynucleotide sequences include oligolabelling, nick translation, end-labelling or PCR amplification using a labelled nucleotide. Alternatively, the coding sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labelled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway, NJ), Promega (Madison, WI), and US Biochemical Corp (Cleveland, OH) supply commercial kits

and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US-A-3817837; US-A-3850752; US-A-3939350; US-A-3996345; US-A-4277437; US-A-4275149 and US-A-4366241. Also, recombinant immunoglobulins may be produced as shown in US-A-4816567.

Additional methods to quantify the expression of a particular molecule include radiolabeling (Melby PC *et al* 1993 J Immunol Methods 159:235-44) or biotinylating (Duplaa C *et al* 1993 Anal Biochem 229-36) nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated. Quantification of multiple samples may be speeded up by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or calorimetric response gives rapid quantification.

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression should be confirmed. For example, if the nucleotide sequence is inserted within a marker gene sequence, recombinant cells containing the same may be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a target coding sequence under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the target as well.

Alternatively, host cells which contain the coding sequence for the target and express the target coding regions may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridisation and protein bioassay or immunoassay techniques which include membrane-based, solution-based, or chip-based technologies for the detection and/or quantification of the nucleic acid or protein.

AGENT

The present invention encompasses the use of kits comprising diagnostic agents and/or therapeutic agents, as well as agents identified using the screening methods described herein.

The agent may be any suitable agent that can act as a modulator of the *RPGR* gene (such as the mutant *RPGR* gene).

The agent can be an amino acid sequence or a chemical derivative thereof. The substance may even be an organic compound or other chemical. The agent may even be a nucleotide sequence - which may be a sense sequence or an anti-sense sequence. The agent may even be an antibody.

Thus, the term "agent" includes, but is not limited to, a compound which may be obtainable from or produced by any suitable source, whether natural or not. The agent may be designed or obtained from a library of compounds which may comprise peptides, as well as other compounds, such as small organic molecules and particularly new lead compounds. By way of example, the agent may be a natural substance, a biological macromolecule, or an extract made from biological materials such as bacteria, fungi, or animal (particularly mammalian) cells or tissues, an organic or an inorganic molecule, a synthetic agent, a semi-synthetic agent, a structural or functional mimetic, a peptide, a peptidomimetics, a derivatised agent, a peptide cleaved from a whole protein, or a peptides synthesised synthetically (such as, by way of example, either using a peptide synthesizer or by recombinant techniques or combinations thereof, a recombinant agent, an antibody, a natural or a non-natural agent, a fusion protein or equivalent thereof and mutants, derivatives or combinations thereof.

As used herein, the term "agent" may be a single entity or it may be a combination of agents.

If the agent is an organic compound then that organic compound may typically comprise one or more hydrocarbyl groups. Here, the term "hydrocarbyl group" means a group comprising at least C and H and may optionally comprise one or more other suitable substituents. Examples of such substituents may include halo-, alkoxy-, nitro-, an alkyl group, a cyclic group etc. In addition to the possibility of the substituents being a cyclic group, a combination of substituents may form a cyclic group. If the hydrocarbyl group comprises more than one C then those carbons need not necessarily be linked to each other. For example, at least two of the carbons may be linked via a suitable element or group. Thus, the hydrocarbyl group may contain hetero atoms. Suitable hetero atoms will be apparent to those skilled in the art and include, for instance, sulphur, nitrogen and oxygen.

The agent may be in the form of a pharmaceutically acceptable salt – such as an acid addition salt or a base salt – or a solvate thereof, including a hydrate thereof. For a review on suitable salts see Berge et al, J. Pharm. Sci., 1977, 66, 1-19.

Suitable acid addition salts are formed from acids which form non-toxic salts and examples are the hydrochloride, hydrobromide, hydroiodide, sulphate, bisulphate, nitrate, phosphate, hydrogen phosphate, acetate, maleate, fumarate, lactate, tartrate, citrate, gluconate, succinate, saccharate, benzoate, methanesulphonate, ethanesulphonate, benzenesulphonate, p-toluenesulphonate and pamoate salts.

Suitable base salts are formed from bases which form non-toxic salts and examples are the sodium, potassium, aluminium, calcium, magnesium, zinc and diethanolamine salts.

A pharmaceutically acceptable salt of an agent of the present invention may be readily prepared by mixing together solutions of the agent and the desired acid or base, as appropriate. The salt may precipitate from solution and be collected by filtration or may be recovered by evaporation of the solvent.

The agent of the present invention may exisit in polymorphic form.

The agent of the present invention may contain one or more asymmetric carbon atoms and therefore exists in two or more stereoisomeric forms. Where an agent contains an alkenyl or alkenylene group, cis (E) and trans (Z) isomerism may also occur. The present invention includes the individual stereoisomers of the agent and, where appropriate, the individual tautomeric forms thereof, together with mixtures thereof.

Separation of diastereoisomers or cis and trans isomers may be achieved by conventional techniques, e.g. by fractional crystallisation, chromatography or H.P.L.C. of a stereoisomeric mixture of the agent or a suitable salt or derivative thereof. An individual enantiomer of the agent may also be prepared from a corresponding optically pure intermediate or by resolution, such as by H.P.L.C. of the corresponding racemate using a suitable chiral support or by fractional crystallisation of the diastereoisomeric salts formed by reaction of the corresponding racemate with a suitable optically active acid or base, as appropriate.

The present invention also includes all suitable isotopic variations of the agent or a pharmaceutically acceptable salt thereof. An isotopic variation of an agent of the present invention or a pharmaceutically acceptable salt thereof is defined as one in which at least one atom is replaced by an atom having the same atomic number but an atomic mass different from the atomic mass usually found in nature. Examples of isotopes that can be incorporated into the agent and pharmaceutically acceptable salts thereof include isotopes of hydrogen, carbon, nitrogen, oxygen, phosphorus, sulphur, fluorine and chlorine such as ^{2}H, ^{3}H, ^{13}C, ^{14}C, ^{15}N, ^{17}O, ^{18}O, ^{31}P, ^{32}P, ^{35}S, ^{18}F and ^{36}Cl, respectively. Certain isotopic variations of the agent and pharmaceutically acceptable salts thereof, for example, those in which a radioactive isotope such as ^{3}H or ^{14}C is incorporated, are useful in drug and/or substrate tissue distribution studies. Tritiated, i.e., ^{3}H, and carbon-14, i.e., ^{14}C, isotopes are particularly preferred for their ease of preparation and detectability. Further, substitution with isotopes such as deuterium, i.e., ^{2}H, may afford certain therapeutic advantages resulting from greater metabolic stability, for example, increased in vivo half-life or reduced dosage

requirements and hence may be preferred in some circumstances. Isotopic variations of the agent of the present invention and pharmaceutically acceptable salts thereof of this invention can generally be prepared by conventional procedures using appropriate isotopic variations of suitable reagents.

It will be appreciated by those skilled in the art that the agent of the present invention may be derived from a prodrug. Examples of prodrugs include entities that have certain protected group(s) and which may not possess pharmacological activity as such, but may, in certain instances, be administered (such as orally or parenterally) and thereafter metabolised in the body to form the agent of the present invention which are pharmacologically active.

It will be further appreciated that certain moieties known as "pro-moieties", for example as described in "Design of Prodrugs" by H. Bundgaard, Elsevier, 1985 (the disclosured of which is hereby incorporated by reference), may be placed on appropriate functionalities of the agents. Such prodrugs are also included within the scope of the invention.

The agent may agonise, antagonise, upregulate, or inhibit a suitable target.

The agent may be a single entity that is capable of exhibiting two or more of these properties. Alternatively, or in addition, the agent can be a combination of agents that are capable of exhibiting one or more of these properties.

Preferably, the agent may selectively agonise, selectively antagonise, selectively upregulate, or selectively inhibit a suitable target.

Preferably, the agent may selectively agonise, selectively antagonise, selectively upregulate, or selectively inhibit a selective, suitable target.

The agent of the present invention may also be capable of displaying one or more other beneficial functional properties.

The agent may be used in combination with one or more other pharmaceutically active agents.

If a combination of active agents are administered, then they may be administered simultaneously, separately or sequentially.

CHEMICAL SYNTHESIS METHODS

If the agent is an organic molecule, then typically the agent of the present invention will be prepared by chemical synthesis techniques.

The agent or target of the present invention or variants, homologues, derivatives, fragments or mimetics thereof may be produced using chemical methods to synthesize the agent in whole or in part. For example, peptides can be synthesized by solid phase techniques, cleaved from the resin, and purified by preparative high performance liquid chromatography (e.g., Creighton (1983) Proteins Structures And Molecular Principles, WH Freeman and Co, New York NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra).

Direct synthesis of the agent or variants, homologues, derivatives, fragments or mimetics thereof can be performed using various solid-phase techniques (Roberge JY et al (1995) Science 269: 202-204) and automated synthesis may be achieved, for example, using the ABI 43 1 A Peptide Synthesizer (Perkin Elmer) in accordance with the instructions provided by the manufacturer. Additionally, the amino acid sequences comprising the agent or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with a sequence from other subunits, or any part thereof, to produce a variant agent.

In an alternative embodiment of the invention, the coding sequence of the agent or variants, homologues, derivatives, fragments or mimetics thereof may be synthesized,

in whole or in part, using chemical methods well known in the art (see Caruthers MH et al (1980) Nuc Acids Res Symp Ser 215-23, Horn T et al (1980) Nuc Acids Res Symp Ser 225-232).

MIMETIC

The present invention also covers the use or identification of mimetics of agents. As used herein, the term "mimetic" relates to any chemical which includes, but is not limited to, a peptide, polypeptide, antibody or other organic chemical which has the same qualitative activity or effect as a reference agent to a target.

CHEMICAL DERIVATIVE

The term "derivative" or "derivatised" as used herein includes chemical modification of an agent. Illustrative of such chemical modifications would be replacement of hydrogen by a halo group, an alkyl group, an acyl group or an amino group.

CHEMICAL MODIFICATION

In one embodiment of the present invention, the agent may be a chemically modified agent.

The chemical modification of an agent of the present invention may either enhance or reduce hydrogen bonding interaction, charge interaction, hydrophobic interaction, Van Der Waals interaction or dipole interaction between the agent and the target sequence.

In one aspect, the identified agent may act as a model (for example, a template) for the development of other compounds.

RECOMBINANT METHODS

Typically the sequences of the present invention are prepared by recombinant DNA techniques.

ANTIBODIES

In one embodiment of the present invention, the agent of the present invention may be an antibody.

Here the antibody may be one that specifically binds to NOIs or the EPs thereof of the present invention but not to corresponding wild type gene or the expression products thereof.

The antibodies may be tested for specific immunoreactivity with an NOI EP and lack of immunoreactivity to the corresponding wild type gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

Antibodies may be produced by standard techniques, such as by immunisation with the substance of the invention or by using a phage display library.

For the purposes of this invention, the term "antibody", unless specified to the contrary, includes but is not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, fragments produced by a Fab expression library, as well as mimetics thereof. Such fragments include fragments of whole antibodies which retain their binding activity for a target substance, Fv, F(ab') and $F(ab')_2$ fragments, as well as single chain antibodies (scFv), fusion proteins and other synthetic proteins which comprise the antigen-binding site of the antibody. Furthermore, the antibodies and fragments thereof may be humanised antibodies. Neutralizing antibodies, i.e., those which inhibit biological activity of the substance polypeptides, are especially preferred for diagnostics and therapeutics.

If polyclonal antibodies are desired, a selected mammal (e.g., mouse, rabbit, goat, horse, etc.) is immunised with an immunogenic polypeptide bearing a epitope(s) obtainable from an identified agent and/or substance of the present invention. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminium hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. BCG (*Bacilli Calmette-Guerin*) and *Corynebacterium parvum* are potentially useful human adjuvants which may be employed if purified the substance polypeptide is administered to immunologically compromised individuals for the purpose of stimulating systemic defence.

Serum from the immunised animal is collected and treated according to known procedures. If serum containing polyclonal antibodies to an epitope obtainable from an identifed agent and/or substance of the present invention contains antibodies to other antigens, the polyclonal antibodies can be purified by immunoaffinity chromatography. Techniques for producing and processing polyclonal antisera are known in the art. In order that such antibodies may be made, the invention also provides polypeptides of the invention or fragments thereof haptenised to another polypeptide for use as immunogens in animals or humans.

Monoclonal antibodies directed against epitopes obtainable from an identifed agent and/or substance of the present invention can also be readily produced by one skilled in the art. The general methodology for making monoclonal antibodies by hybridomas is well known. Immortal antibody-producing cell lines can be created by cell fusion, and also by other techniques such as direct transformation of B lymphocytes with oncogenic DNA, or transfection with Epstein-Barr virus. Panels of monoclonal antibodies produced against orbit epitopes can be screened for various properties; i.e., for isotype and epitope affinity.

Monoclonal antibodies to the substance and/or identified agent of the present invention may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique originally described by Koehler and Milstein (1975 Nature 256:495-497), the human B-cell hybridoma technique (Kosbor *et al* (1983) Immunol Today 4:72; Cote *et al* (1983) Proc Natl Acad Sci 80:2026-2030) and the EBV-hybridoma technique (Cole *et al* (1985) Monoclonal Antibodies and Cancer Therapy, Alan R Liss Inc, pp 77-96). In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison *et al* (1984) Proc Natl Acad Sci 81:6851-6855; Neuberger *et al* (1984) Nature 312:604-608; Takeda *et al* (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies (US Patent No. 4,946,779) can be adapted to produce the substance specific single chain antibodies.

Antibodies, both monoclonal and polyclonal, which are directed against epitopes obtainable from an identifed agent and/or substance of the present invention are particularly useful in diagnosis, and those which are neutralising are useful in passive immunotherapy. Monoclonal antibodies, in particular, may be used to raise anti-idiotype antibodies. Anti-idiotype antibodies are immunoglobulins which carry an "internal image" of the substance and/or agent against which protection is desired. Techniques for raising anti-idiotype antibodies are known in the art. These anti-idiotype antibodies may also be useful in therapy.

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in Orlandi *et al* (1989, Proc Natl Acad Sci 86: 3833-3837), and Winter G and Milstein C (1991; Nature 349:293-299).

Antibody fragments which contain specific binding sites for the substance may also be generated. For example, such fragments include, but are not limited to, the $F(ab')_2$

fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the $F(ab')_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse WD *et al* (1989) Science 256:1275-128 1).

TREATMENT

It is to be appreciated that all references herein to treatment include curative, palliative and prophylactic treatment.

The therapeutic regime of the present invention may be tailored to the needs of the individual being treated and exposure to adverse side effects minimised. The present invention can therefore be utilised to identify which individuals would be most likely to benefit from, for example, gene therapies. Gene therapy techniques include but are not limited to techniques which replace a faulty gene, such as a *RPGR* gene (such as the mutant *RPGR* gene) and/or which downregulate expression of the gene and/or function of the gene product.

MODULATORS

The present invention also encompass modulators of the *RPGR* gene (such as the mutant *RPGR* gene), such as those identified using the assay method(s) of the presen invention. Examples of modulators of a *RPGR* gene (such as the mutant *RPGR* gene) of the present invention include but are not limited to compounds and substances that affect the expression of the gene as well as the activity and/or amount of the expressed gene product. Typical modulators suitable for use in the invention include *RPGR* gene (such as the mutant *RPGR* gene) agonists and/or antagonists - either of a *RPGR* gene (such as the mutant *RPGR* gene) or of the *RPGR* gene (such as the mutant *RPGR* gene) product; large and small molecular weight inhibitors of *RPGR* gene (such as the mutant *RPGR* gene) expression or function; inducers and suppressors of a *RPGR* gene(such as the mutant *RPGR* gene); antisense sequences - including antisense

oligonucleotides to a *RPGR* gene or transcript; antibodies; *RPGR* gene (such as the mutant *RPGR* gene) product binding proteins - including dominant negative versions of the *RPGR* gene (such as the mutant *RPGR* gene) product that may be involved in oligomerisation; and *RPGR* gene (such as the mutant *RPGR* gene) product kinases; or fragments, variants and derivatives thereof.

An example of a modulator according to the invention is an antisense oligonucleotide that binds to and prevents or reduces transcription of *RPGR* (such as the mutant *RPGR* gene) mRNA. This modulator may be used to reduce the activity and/or amount of a *RPGR* gene (such as the mutant *RPGR* gene) EP.

As used herein, the term "antisense" is used to refer to a nucleic acid strand that is complementary to the "sense" strand. The designation (-) is sometimes used in reference to the antisense strand, with the designation (+) sometimes used in reference to the sense (or positive) strand. An antisense nucleic acid to the *RPGR* NOIs of the present invention may be produced by any method including the synthesis of the *RPGR* NOIs in a reverse orientation to a promoter which permits the synthesis of the coding strand. This transcribed strand may combine with, for example, a natural mRNA produced by a cell to form a duplex. These duplexes may then block either the further transcription of the mRNA and/or its translation.

Another suitable modulator is the EP itself, which can be administered by injection and is used to increase the activity and/or amount of the *RPGR* EP.

A further embodiment of the present invention provides for a method of preventing and/or treating disease comprising administering a modulator of a *RPGR* NOI.

ADMINISTRATION OF MODULATORS

The modulators of the *RPGR* gene (such as the mutant *RPGR* gene) of the present invention may be suitable for a number of disease states and conditions. Such modulators may be suitable for either prophylactic administration or after a disease has

been diagnosed. The route of administration is suitably chosen according to the disease or condition to be treated, however, typical routes of administration of the modulator of the present invention include but are not limited to oral, rectal, intravenous, parenteral, intramuscular and sub-cutaneous routes. The invention also provides for *RPGR* modulators to be administered either as DNA or RNA and thus as a form of gene therapy, or as proteins. The modulators may be delivered into cells directly by means including but not limited to liposomes, viral vectors and coated particles (gene gun).

The modulators of the invention may be suitable for treatment of a range of diseases and conditions. In some individuals a combination of diseases may be present or predicted wherein in others only one is diagnosed.

EXAMPLES

The invention will now be further described only by way of example in which reference is made to the following Figures:

FIGURES

Figure 1 which shows a series of images;
Figure 2 which shows sequences;
Figure 3 which shows a series of images;
Figure 4 which shows a series of sequences; and
Figure 5 which shows a series of images.

In more detail:

Fig. 1 - Alternative splicing of human *RPGR*.

Total RNA was prepared from human tissues and cell lines, and specific transcripts were amplified by RT-PCR. The region analysed is represented schematically on the right. An exon is shown as a numbered box. The position and orientation of the

primers used for RT-PCR is indicated with arrowheads. When hemi-nesting was required (panels *b* and *c*), all three primers are shown. The identity of all relevant products was verified by direct sequencing. The non-specific products generated in one of the experiments (panel *b*) are marked with a star. Results shown here are for testis (T), ARPE-19 cell line (A), Weri-Rb-1 retinoblastoma cell line (W), Y79 retinoblastoma cell line (Y), skeletal muscle (M), brain (B), liver (L), kidney (K), heart (H), lung (l), pancreas (P), spleen (S), adrenal (a), and retina (R). In some panels a molecular weight marker is shown (X). A fragment of the *GAPDH* mRNA was amplified in a control reaction (panel *h*), with commercial primers (Clontech, Palo Alto, CA). Exons 15b1 and 15b2 are two overlapping exons derived from intron 15, using alternative acceptor splice sites and the same donor site. Their inclusion is predicted to result in premature termination of translation. Nested PCR was necessary to detect these exons, suggesting a low level of expression.

Fig. 2 - Nucleotide and deduced amino-acid sequence of human exon ORF15.

ORF15 is a novel 3' terminal exon, spliced to exon 14. It shares its acceptor site with exon 15. Exon 15 is highlighted in grey. The consensus polyadenylation signal is underlined with a double line. The positions of the disease causing sequence alterations found in XLRP patients are highlighted in black. Potentially benign or polymorphic sequence variants are underlined. They are substitutions (g.ORF15+470G/A, g.ORF15+1466C/T), or in frame deletions and duplications (g.ORF15+914_916delGGA,g.ORF15+1307_1318del12, g.ORF15+1321_1332del12, g.ORF15+1067_1087dup21, g.ORF15+1165_1185dup21). All deletions and duplications were found on control chromosomes. Although the 15nt deletion of patient 18 was not found in controls, it is probably not disease-causing since he also carried a g.ORF15887G>T, which generates a premature nonsense codon.

Fig. 3 - Mutations in ORF15 in XLRP patients.

a, ORF15 was initially screened for mutations with SSCA in XLRP patients, and changes were found in 16 cases. A representative SSCA result is shown. *b*, The

nucleotide changes underlying the altered SSCA pattern were determined by direct sequencing. Representative results are shown. *c*, The most repetitive region of ORF15 could not be analysed accurately by SSCA, and was sequenced directly from PCR products. Mutations in another 12 patients were found, some of which are shown here. *d*, Identical mutations were found in some XLRP families. Analysis of 8 intragenic polymorphic positions is shown for 4 of these mutations. Each deduced haplotype is shown with a different colour. These results suggest that each mutation has occurred at least twice through recurrent mutation *e*, The distribution of published *RPGR* mutations (exons shaded in grey) or in the present series (dots) are shown in relation to the different *RPGR* alternative transcripts, suggesting that the transcript consisting of exons 1-14 and ORF15 is important in XLRP and for *RPGR* function in the retina. Exons 16-19, in which no mutations have been found, are not used in this transcript. An alternative explanation for the lack of mutations in exons 16-19 is the inclusion of an alternative exon coding for a stop codon (15a, and 15b1/2 not shown), and is not supported by the present results.

Fig. 4 - Conservation of *RPGR* ORF14 and ORF15.

Like human *RPGR* (panel *a*), the mouse (*b*), bovine (*c*) and *Fugu (d) RPGR* genes have a large open reading frame in the region corresponding to ORF14 and ORF15. The predicted proteins are shown. The second two thirds of each sequence is very repetitive, and has an unusually high content of glutamic and/or aspartic acid in all species (highlighted in light grey), and of glycine in mammals (dark grey). *e*, Repetitive sequence in human *RPGR* exon ORF15. The most repetitive sequence of ORF15 (nt 705-1406) consists of 27 imperfect direct repeats of 15-33 nucleotides, coding for the consensus peptide sequence E(1-5)GEGEGE. f, the C-terminus of ORF15 is well conserved through evolution. Residues conserved in all species are highlighted in black, those conserved in at least three species in grey. Dots indicate gaps introduced to optimise the alignments

Fig. 5 - Expression of ORF15 and ORF14/15 in mouse and cow.

Total RNA of various tissues was prepared, and specific transcripts were amplified with RT-PCR. The region analysed is represented schematically on the right. An exon is shown as a numbered box, exons 12-15 are highlighted in grey, the position and orientation of the primers used for RT-PCR is indicated with arrowheads. The identity of all relevant products was verified by direct sequencing. Results shown are for 661 cells (6), ovary (O), testis (T) skeletal muscle (M), brain (B), liver (L), kidney (K), heart (H), lung (l), pancreas (P), spleen (S), adrenal medulla (a), adrenal cortex (C), eye (E) and retina (R). In some panels a molecular weight marker is shown (X). A fragment of the *GAPDH* mRNA was amplified in a control reaction. *a*, ORF15 is preferentially expressed in the mouse retina. A variant where intron 14 was retained (exon ORF14/15) was found exclusively in the eye/retina. *b*, In bovine tissues the ORF14/15 transcript is detected at highest levels in the retina, and at lower level in testis. *c*, Comparison of intron 14 acceptor site in human, mouse and cow. Intron 14 appears to be retained in bovine tissues because its acceptor site is not conserved.

MATERIALS AND METHODS

Human XLRP samples.

The present series of XLRP patients consists of 47 families diagnosed as described previously[1], with 87% of families coming from the UK or Ireland. Patient 55 was diagnosed as suffering from a probable X-linked cone dystrophy (see Table 1).

Human cosmid sequencing.

The isolation and sequencing of cosmids E75, Y95, C4, Y94, Y91 and M32 was carried out as described[1]. The genomic sequence of the remaining two gaps between cosmids E75 and C4 (3.7 kb) and between cosmids Y94 and M32 (6.7 kb) were filled by long-range PCR amplication of Y95 and Y91 respectively. Each fragment was subcloned into the Expand™ Cloning vector (Boehringer) and sequenced after subcloning into M13. The sequence contains two earlier database submissions, X94768 and X94767 (ref. 15). The latter represent 23 kb of sequence from the 3'

end of the RPGR gene, including 3.5 kb of intron 15, and extend in a 3' direction to exon 19, with a further 12 kb downstream of this exon.

***Fugu rubripes*, mouse and bovine *RPGR* sequence.**

A cosmid containing the *Fugu RPGR* homologue was obtained from the UK Human Genome Mapping Project Resource Centre (HGMP-RC; Hinxton, Cambridge, UK), and was partially sequenced directly using primer walking (R. Vervoort, manuscript in preparation). A mouse PAC containing the entire *mRpgr* gene was isolated, and used to amplify the exon 14-16 region with XL-PCR, with Expand™ Long Template PCR System (Boehringer). PCR products were cloned in pCR®-XL-TOPO® (Invitrogen) and sequenced directly with primer walking at Oswel DNA Sequencing (Southampton, UK). The sequence of mouse ORF14/15 was verified in 2 independent shorter PCR products, amplified with different primers. The sequence of bovine *RPGR* cDNA was determined (R. Vervoort, manuscript in preparation) and used to design oligonucleotides for amplification of bovine ORF14/15 from a Universal GenomeWalker™ library (Clontech), constructed from bovine genomic DNA, according to the manufacturers instructions. Specific products were cloned in pCR®-TOPO (Invitrogen) and sequenced using primer walking at Oswel DNA Sequencing (Southampton, UK).

Sequence analysis.

The gene sequence was analysed on both strands with on-line programs, including CENSOR (ref. 16), MZEF (ref. 17), GRAIL II (ref. 18), GENSCAN (ref. 19), and GENIE (ref. 20). In addition, the sequence was submitted to NIX at HGMP (ref. 21, *http://www.hgmp.mrc.ac.uk*). NIX is a WWW tool to view the results of running several different DNA analysis programs, including GRAIL, Fex, HMMgene, MZEF, GENSCAN, Genemark, Genefinder, FGene and BLAST, and to screen sequence databases including dbEST, SwissProt, RepeatMasker and tRNAscan. Proteins were scanned for motifs using on-line software, including PIX at HGMP and facilities at GenomeNet (*http://www.motif.genome.ad.jp/*).

RT-PCR and 3' RACE analyses.

Total RNA was prepared from cultured cells or tissues using TRIZOL™ Reagent (Life Technologies). Reverse transcription and PCR were performed with the GeneAmp reagents (Perkin-Elmer) according to the manufacturers instructions; for a standard 100 µl reaction 1 µg of total RNA was used, and the reverse transcription was generally primed with random hexamers. RACE was performed using the Not I-$d(T)_{18}$ primer (Pharmacia) as a primer for reverse transcription; subsequent PCR amplification was with primer Not27 antisense primer, consisting of the anchor portion of the Not I-$d(T)_{18}$ primer. Products were sequenced directly with ABI Prism™ dRhodamine Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems) and analysed on an ABI Prism 377 Automated Sequencer.

Mutation analysis.

SSCA of exons 1-19, and sequencing of corresponding PCR products was performed as described previously[1]. Exons 15a and 15b1/2 were analysed in essentially the same way, as was SSCA of 12 different fragments covering ORF14 and part of ORF15 (nt 1-752 and 1274-1718). ORF15 nt 753-1273 could not be analysed accurately with SSCA, and was sequenced directly from PCR products, with ABI Prism™ dRhodamine Terminator Cycle Sequencing Ready Reaction kit (PE Applied Biosystems), and analysed on an ABI Prism 377 Automated Sequencer.

Cell lines.

The human retinoblastoma cell lines WERI-Rb-1 and Y79 were obtained from the ATCC. Cell culture was according to the instructions provided by ATCC. The human RPE cell line ARPE19 was kindly provided by Dr. L. Hjelmland[22]. The mouse retina 661W cell line was kindly provided by Dr. M. Al-Ubaidi.

EXPERIMENTAL RESULTS AND DISCUSSION

The present invention is based on the finding of a mutational hot spot within a novel *RPGR* exon in X-linked retinitis pigmentosa.

A gene called *RPGR* was previously identified in the *RP3* region of Xp21.1 and shown to be mutated in 10-20% of patients with the progressive retinal degeneration X-linked retinitis pigmentosa (XLRP)[1-2]. The number of mutations was less than the 70-75% expected from linkage studies[3-6]. Mutations in the *RP2* gene in Xp11.3 were found in a further 10-20% of XLRP patients, as predicted from linkage studies[7-8]. Since the missing mutations may reside in undiscovered exons of the *RPGR* gene, a 172 kb region containing the entire *RPGR* gene was sequenced. Analysis of the sequence disclosed a novel 3' terminal exon, which was mutated in 60% of XLRP patients. The exon codes for 567 amino-acids, with a repetitive domain rich in glutamic acid residues. The sequence is conserved in the mouse, bovine and *Fugu* genes. It is preferentially expressed in mouse and bovine retina, further supporting its importance for retinal function. These results suggest that mutations in the *RPGR* gene are the only cause of XLRP/*RP3* and account for the disease in over 70% of XLRP patients and an estimated 11% of all retinitis pigmentosa patients.

To identify novel sequences necessary for the function of *RPGR*, shotgun sequencing of six overlapping cosmids spanning 172 kb containing the sequence between exon 1 of the *ETX1/SRPX* gene and exon 5 of the *OTC* gene was carried out[1,9-10]. The sequence was first analysed with exon prediction programs. Reverse transcription-polymerase chain reaction (RT-PCR) analysis of total RNA from eleven adult tissues and three cell lines was carried out to verify the predictions. These experiments demonstrated the expression of five novel exons: ORF14, 15b1, 15b2, 15a, and ORF15 (Fig. 1). ORF14 and ORF15 are parts of a 2.2 kb open reading frame, predicted to be an exon by most of the programs. Exons 15b1/2 and 15a are small exons in intron 15. Besides the 19-exon mRNA reported previously[1], two transcripts were relatively abundant in retina: one was widely expressed and lacked exons 14-15 (Fig. 1*g*), another was preferentially expressed in the retina and contained exon 15a

(Fig. 1*a,f*)[11]. Specific primers were necessary within exons 15b1/b2 and ORF14 to amplify the corresponding transcripts, indicating their low relative abundance.

ORF14 is a large internal exon generated by retention of intron 14. The 2.2 kb open reading frame in this region extends into intron 15, and evidence for expression of this sequence was found in cDNA (Fig. 1*e*): the novel exon ORF15 shares its 3'splice site with exon 15, and is spliced to exon 14. It was found in all tissues examined, with the most prominent bands in retina and retinal cell lines. No specific products were obtained when we tried to link ORF15 to exons 16-19, suggesting that ORF15 is an alternative 3' terminal exon. 3' RACE of retina cDNA confirmed the presence of a polyadenylation tract at ORF15 position 2834, preceded by a polyadenylation signal at position 2818 (Fig. 2).

To address the functional significance of the novel exons ORF14, ORF15, 15b1, 15b2 and 15a, these regions were screened for mutations, initially by single stranded conformational analysis (SSCA) analysis in 47 XLRP patients. No mutations were found in exons 15a, 15b1 and 15b2. Except for a known polymorphism (1765G/A) in one of the patients, no sequence alterations were detected in ORF14. However, PCR products corresponding to fragments of ORF15 showed aberrant migration on SSCA gels in 16 patients. The underlying sequence alterations consisted of seven different 1, 2 or 4 nucleotide (nt) deletions and a large duplication of 73 nt. The most repetitive region of ORF15, which could not be analysed accurately by SSCA, was examined by direct sequencing of PCR products spanning the coding region of ORF15. A further five different 1, 2 and 5 nt deletions, one 1 nt insertion and three substitutions leading to a nonsense mutation were identified in 12 patients (Table 1, Fig. 3). A total of 28 patients had presumed mutations in ORF15, each of which lead to premature termination of translation. This is four times as many as the 6 mutations found in exons 1-14 (Table 1). None of these mutations has been detected in 150 control chromosomes. Mutation analysis of ORF15 also revealed several changes thought to be benign sequence variants (Fig. 2), including two nucleotide substitutions and five different in-frame rearrangements of 3, 12 or 21 nt, which were found in control chromosomes.

TABLE 1

Sample	Family	Mutation name	Predicted effect
3	5	g.ORF15+1239_1243del5	frameshift
4	7	g.ORF15+1344delG	frameshift
5	8	g.ORF15+1297_1298delAG	frameshift
6	9	g.ORF15+689_692del4	frameshift
7	10	g.ORF15+652_653delAG	frameshift
11	6	g.ORF15+977_978delGG	frameshift
13	17	g.ORF15+673_674delAG	frameshift
14	19	g.ORF15+483_484delGA	frameshift
16	3	g.ORF15+1239_1243del5	frameshift
18	25	g.ORF15+887G>T	p.ORF15E299X
		g.ORF15+694_708del15	
19	26	g.ORF15+673_674delAG	frameshift
20	27	g.ORF15+652_653delAG	frameshift
21	28	g.ORF15+1239_1243del5	frameshift
23	35	g.ORF15+769delA	frameshift
24	38	g.ORF15+391_463dup73	frameshift
25	39	g.ORF15+652_653delAG	frameshift
27	43	g.ORF15+492G>T	p.ORF15E164X
29	47	353C>A	H98Q
31	50	640G>A	W194X
32	51	g.ORF15+804G>T	p.ORF15E268X
37	57	g.ORF15+1141delA	frameshift
39	60	g.ORF15+689_692del4	frameshift

40	61	g.ORF15+673_674delAG	frameshift
42	62	474G>T	E139X
43	76	g.ORF15+1010_1011delGG	frameshift
46	73	g.ORF15+647_648delAA	frameshift
47	68	g.ORF15+689_692del4	frameshift
48	22	g.ORF15+673_674delAG	frameshift
50	63	523+1G>T	frameshift
52	70	807T>C	C250R
55	78	g.ORF15+689_692del4	frameshift
56	79	g.ORF15+872_873insA	frameshift
57	80	g.ORF15+977_978delGG	frameshift
58	81	g.EX8+57_IVS8+1317del1374	c.838_993del

Table 1 presents the mutation analysis of *RPGR* in XLRP families. Affected male probands from 47 XLRP families were screened for mutations in the *RPGR* gene. Six mutations were found in the RCC1 domain. Four of these have been reported earlier[1,23], the fifth is a G>T substitution in exon 5 resulting in a E139X nonsense mutation. The sixth mutation is a 1374 nt deletion, leading to skipping of exon 8 at the mRNA level (R. Vervoort, unpublished data). The majority of patients have a mutation in ORF15; numbering of these mutations refers to the position in the exon (see also Fig. 2). Patient 55 was diagnosed as a "probable" X-linked cone dystrophy and was found to have an *RPGR* mutation.

No functional protein motifs were found in ORF14 and ORF15. Because of its repetitive sequence and high glutamic acid and glycine content, the predicted ORF15 protein is highly unusual. Although there are many proteins with short acidic domains in the databases, the only close resemblance is a viral protein of unknown function, VG48_HSVSA (accession no. Q01033). To assess the functional importance of ORF14 and ORF15, and as part of a study of the evolution of *RPGR*

(R. Vervoort, unpublished data), the corresponding region in mouse, cow and *Fugu rubripes* was sequenced. The gene in all four species contains a large open reading frame with a purine-rich, repetitive 3' half (Fig. 4*a-d*). The predicted ORF14 region is well conserved in mammals, but not in *Fugu* (Fig. 4*e*). All ORF15 proteins are predicted to be rich in glutamic and/or aspartic acid residues, alternating with glycine in the mammalian proteins (Fig. 4*a-d*). Despite this highly similar amino acid content, the sequences have diverged considerably at the primary sequence level. The C-terminus of ORF15 is well conserved in all species, including *Fugu* (Fig 4*f*).

The expression of mouse and bovine ORF15 was analysed by RT-PCR (Fig. 5). This analysis revealed a further novel transcript in which ORF14 and ORF15 sequences are used together as an exon called ORF14/15, spliced to exon 13. Expression of ORF14/15 was limited to the retina (mouse) or testis and retina (bovine). Mouse ORF15 alone was expressed in a wide range of tissues but preferentially in the retina, as with human ORF15. ORF15 alone was not found in bovine tissues, presumably because the acceptor splice site of intron 14 is not conserved (Fig. 5*c*). The ORF14/15 splice variant was not detected in human retina and the lack of mutations in intron 14 suggests that it is not important in human retinal disease. The clustering of mutations in terminal exon ORF15, the presence of published disease-causing mutations in exons 1-14 and lack of reported mutations in exons 16-19 all support the view that the transcript with exons 1-14 and ORF15 is responsible for the retinal degeneration in XLRP (Fig 3*e*).

The high frequency of mutations in the terminal exon ORF15 (17 different mutations in 1 kb) compared with other parts of the same *RPGR* transcript (6 mutations in 1.6 kb), suggests that it is a mutation hot spot. Five different mutations were found to exist on at least two different haplotypes suggesting recurrent mutation (Fig 3*d*), which increases the frequency to at least 24 independent mutations in 1 kb. The high mutability of ORF15 may be related to its unusual nucleotide composition and/or the repetitive nature of the sequence. The sequence of ORF15 is purine-rich: all mutations occur in a 1061 nt coding strand with only 27 (2.5 %) pyrimidines. This type of sequence may adopt unusual non B-DNA conformations, including triplex

structures, which are associated with reduced fidelity of replication[12]. A 6 nt motif similar to DNA polymerase α arrest sites has been found near deletion hotspots in other human genes[13]. The ORF15 sequence contains numerous potential polymerase arrest sites suggesting that arrest may occur during replication leading to slipped strand mispairing events, since many ORF15 mutations involve direct repeats[13-14].

In summary, these results suggest that a transcript consisting of the first 14 exons and the novel 3' terminal exon, ORF15, is necessary for the normal function of the *RPGR* gene in the human retina. This transcript is affected by all RP3 mutations documented, of which 20% are in exons 1-14, and 80% are in the repetitive purine-rich sequence of ORF15 (Fig 3*e*). In this series, *RPGR* mutations have been found in 72% of XLRP patients, suggesting that at least 11% of all RP referrals may be accounted for by this locus[1]. *RPGR* exons 2-10 code for a domain homologous to the RCC1 protein, a guanine nucleotide exchange factor for the small GTPase Ran. The identification of the new ORF15 domain is an important step towards a better understanding of the role of *RPGR* in health and disease.

EXPERIMENTAL SUMMARY

The diagnosis of X-linked retinitis pigmentosa (XLRP) is difficult since there are no clinical means of reliably distinguishing it from other forms of retinitis pigmentosa (RP). This condition is clinically one of the most severe forms of RP, with onset in the first decade of life and severe visual impairment by the fourth decade. The diagnosis has major implications for families since a female carrier will have a 1 in 2 chance of having a son with severe disease. There is therefore considerable demand for an efficient diagnostic test. XLRP affects 16-33% of all RP patients and genetic mapping studies suggested that about 75% of families mapped to chromosomal region Xp21.1. A gene was isolated in from Xp21.1 in 1996 which was found to be responsible for mutations in 15-20% of XLRP patients (Meindl et al., 1996), which was later confirmed by others.

We have now discovered that a large proportion of the remaining XLRP mutations are contained within a 1 kilobase region of a novel RPGR exon, ORF15, with unusual sequence characteristics. This region was found to contain a high proportion of rearrangements and point mutations (mutation hot spot) that are predicted to interrupt the normal protein reading frame (frameshifts), with consequent loss of function. Since it is now possible to detect disease-causing mutations in a relatively large proportion (10-20%) of all RP patients, we wish to patent the application of mutation analysis of the RPGR exon ORF15 for diagnostic purposes.

At present, XLRP patients comprise a large subgroup (16-33%) of all RP patients for whom there is presently no efficient means of mutation detection, since previously known RPGR exons only detected mutations in a small fraction (15-20%) of XLRP patients. RP is clinically very heterogeneous with at least 30 and possibly double this number of genes causing the disease in different families. The majority of these genes affect only a small proportion (1-2%) of patients. Mutation analysis of the RPGR ORF15 exon will for the first time provide a reliable diagnosis in a large proportion of XLRP patients. About 80% of all RPGR mutations lie within this relatively small 1kb region of exon ORF15.

The application of DNA-based diagnostic tests has considerable potential in severe conditions in which the test provides a clear-cut prediction. In this case, the proposed test predicts with very high certainty the presence or absence of a gene causing a severe blinding disease.

There is a large demand for such diagnostic test, since the presence of an XLRP mutation carries important implications for reproductive risks, particularly for potential carrier females. Carrier females are generally mildly affected but affected males severely so. In the future, it is likely that therapy will become available by gene replacement or generic means (e.g. neuroprotective factors) applied locally to the retina, which is a relatively accessible tissue. The prevalence of RP is in the region of 1 in 3000 worldwide, of whom RPGR mutations are likely to account for 11-23%. Over 50% of all RP patients are of unknown genetic type, so many of the

male patients from this group may be advised to have an RPGR mutation test, in view of the implications for their families. The total proportion of RP patients requiring such a test may therefore reach 30-40%, which is 1/7,500-1/10,000, or about 6,000 patients in the UK and 25,000 patients in USA.

In addition, having ascertained an RPGR mutation, several other family members are likely to be offered testing. In total, we estimate that a maximum of 1/5,000 of the general population might require testing, of which perhaps one-third might take advantage of it (1/15,000). The current estimated population of countries with established market economies is 840 million, from which some 56,000 people might wish to take advantage of this test.

The test we provide is relatively straightforward, involving the techniques of polymerase chain reaction (PCR) and DNA sequencing, which are widely used and robust. The target sequence is also short (1 kb) making it amenable to a simple test procedure on a DNA sample from mouth wash or blood. The interpretation of test results are general unambiguous since all ORF15 mutations to date and 85-90% of all RPGR mutations are predicted to result in an altered translational reading frame and loss-of-function.

INVENTION SUMMARY

In summation, some broad aspects of the present invention will now be described by way of numbered paragraphs:

1. A method of diagnosis for a disease or a predisposition to a disease associated with a disease causing mutation(s) in a *RPGR* gene;
wherein the method comprises:
genotyping a *RPGR* gene; and
determining whether the genotype comprises a disease causing mutation(s).

Preferably, said disease causing mutation(s) is/are located towards the 3' end of the *RPGR* gene.

Preferably, said disease causing mutation(s) is/are located in an exon located towards the 3' end of the *RPGR* gene.

2. A kit for the diagnosis of a disease or a predisposition to disease; wherein the kit comprises:
means for genotyping a *RPGR* gene; and
reference means for determining whether the genotype comprises a disease causing mutation(s).

Preferably, said disease causing mutation(s) is/are located towards the 3' end of the *RPGR* gene.

Preferably, said disease causing mutation(s) is/are located in an exon located towards the 3' end of the *RPGR* gene.

3. A method of preventing and/or treating a disease or a predisposition to a disease associated with a disease causing mutation(s) in a *RPGR* gene; wherein the method comprises:
genotyping a *RPGR* gene;
determining the presence of a disease causing mutation(s) in the *RPGR* gene; and
applying a treatment in order to prevent, delay, reduce or treat the disease or the predisposition to the disease if said *RPGR* gene comprises said disease causing mutation(s).

Preferably, said disease causing mutation(s) is/are located towards the 3' end of the *RPGR* gene.

Preferably, said disease causing mutation(s) is/are located in an exon located towards the 3' end of the *RPGR* gene.

4. An assay method for identifying an agent capable of modulating a *RPGR* gene or the expression product thereof wherein the assay method comprises:
contacting the agent with a mutant *RPGR* gene or fragment thereof or the expression product thereof;
determining whether the agent modulates the gene or the expression product thereof.

Preferably, said disease causing mutation(s) is/are located towards the 3' end of the *RPGR* gene.

Preferably, said disease causing mutation(s) is/are located in an exon located towards the 3' end of the *RPGR* gene.

5. A process comprising the steps of:
performing the assay according to paragraph 4;
identifying one or more agents capable of modulating the gene or expression product thereof; and
preparing a quantity of one or more of the identified agents.

6. A process comprising the steps of:
performing the assay according to paragraph 4 or paragraph 5;
identifying one or more agents capable of modulating the gene or the expression product thereof; and
preparing a pharmaceutical composition comprising one or more of the identified agents.

7. A process comprising the steps of:
performing the assay according to paragraph 4 or paragraph 5;
identifying one or more agents capable of modulating the gene or the expression product thereof;

modifying one or more of the identified agents; and

preparing a pharmaceutical composition comprising one or more of the modified agents.

8. An agent identified or modified by the process according to any one of paragraphs 4 to 7.

9. A pharmaceutical composition comprising an agent according to paragraph 8 and a pharmaceutically acceptable carrier, diluent, excipient or adjuvant or any combination thereof.

10. A method of preventing and/or treating disease associated with a mutant *RPGR* gene comprising administering an agent according to paragraph 8 or a pharmaceutical according to paragraph 9 wherein said agent or said pharmaceutical is capable of modulating said mutant *RPGR* gene or expression product thereof to cause a beneficial therapeutic effect.

11. A kit according to paragraph 2 wherein the kit additionally comprises an agent according to paragraph 8 or a pharmaceutical according to paragraph 9; wherein said agent or said pharmaceutical is capable of modulating and/or preventing and/or treating a disease associated with a mutant *RPGR* gene.

12. A mutant *RPGR* gene.

13. A nucleotide capable of selectively hybridising to a mutant RPGR gene and not the wild-type *RPGR* gene.

14. A mutant RPGR protein.

Likewise, some preferred aspects of the present invention will now be described by way of numbered paragraphs:

1. A method of diagnosis for a disease or a predisposition to a disease associated with a disease causing mutation(s) in a *RPGR* gene;
wherein the method comprises:
genotyping a *RPGR* gene; and
determining whether the genotype comprises a disease causing mutation(s);
wherein said disease causing mutation(s) is present within ORF15 of the *RPGR* gene.

2. A method according to paragraph 1 wherein said disease causing mutation(s) is present within SEQ ID No. 2.

3. A kit for the diagnosis of a disease or a predisposition to disease; wherein the kit comprises:
means for genotyping a *RPGR* gene; and
reference means for determining whether the genotype comprises a disease causing mutation(s);
wherein said disease causing mutation(s) is present within ORF15 of the *RPGR* gene.

4. A kit according to paragraph 3 wherein said risk genotype is present within SEQ ID No. 2.

5. A nucleotide sequence comprising ORF15 of the *RPGR* gene or a variant, homologue, derivative or fragment thereof, but wherein said nucleotide sequence is not, or is not present within, the wild-type *RPGR* gene.

Preferably said sequence comprises disease causing mutation(s).

6. A nucleotide sequence comprising SEQ ID No. 2 or a variant, homologue, derivative or fragment thereof, but wherein said nucleotide sequence is not present within the wild-type *RPGR* gene.

Preferably said sequence comprises disease causing mutation(s).

7. A construct comprising the sequence of paragraph 5 or paragraph 6.

8. A vector comprising the sequence of paragraph 5 or paragraph 6.

9. A plasmid comprising the sequence of paragraph 5 or paragraph 6.

10. A host cell comprising the sequence of paragraph 5 or paragraph 6.

11. An amino acid sequence encodable by the sequence of paragraph 5 or paragraph 6.

12. A mutant *RPGR* gene, wherein said gene has a mutation in at least ORF15.

13. A nucleotide capable of selectively hybridising to a mutant RPGR gene and not the wild-type *RPGR* gene, wherein said gene has a mutation in at least ORF15.

14. A mutant RPGR protein, wherein said protein has a mutation as a result of a mutation in at least ORF15 of the *RPGR* gene.

15. A method of preventing and/or treating a disease or a predisposition to a disease associated with a disease causing mutation(s) in a *RPGR* gene; wherein the method comprises:
genotyping a *RPGR* gene;
determining the presence of a disease causing mutation(s) in the *RPGR* gene; and
applying a treatment in order to prevent, delay, reduce or treat the disease or the predisposition to the disease if said *RPGR* gene comprises said disease causing mutation(s);
wherein said disease causing mutation(s) is present within ORF15 of the *RPGR* gene.

16. A method according to paragraph 15 wherein said disease causing mutation(s) is present within SEQ ID No. 2.

17. A process for preparing an isolated protein encodable by a nucleotide sequence according to paragraph 5 or paragraph 6, said process comprising expressing a nucleotide sequence according to paragraph 5 or paragraph 6 and optionally isolating and purifying the protein.

18. A protein produced by the process according to paragraph 17.

19. An assay method for identifying an agent capable of modulating a *RPGR* gene or the expression product thereof wherein the assay method comprises:
contacting the agent with the nucleotide sequence according to paragraph 5 or paragraph 6 or the expression product thereof;
determining whether the agent modulates the sequence or the expression product thereof.

20. A process comprising the steps of:
performing the assay according to paragraph 19;
identifying one or more agents capable of modulating nucleotide sequence according to paragraph 5 or paragraph 6; and
preparing a quantity of one or more of the identified agents.

21. A process comprising the steps of:
performing the assay according to paragraph 19 or paragraph 20;
identifying one or more agents capable of modulating a nucleotide sequence according to paragraph 5 or paragraph 6 or the expression product thereof; and
preparing a pharmaceutical composition comprising one or more of the identified agents.

22. A process comprising the steps of:
performing the assay according to paragraph 19 or paragraph 21;
identifying one or more agents capable of modulating a nucleotide sequence according to paragraph 5 or paragraph 6 or the expression product thereof;
modifying one or more of the identified agents; and

preparing a pharmaceutical composition comprising one or more of the modified agents.

23. An agent identified or modified by the process according to any one of paragraphs 19 to 21.

24. A pharmaceutical composition comprising an agent according to paragraph 23 and a pharmaceutically acceptable carrier, diluent, excipient or adjuvant or any combination thereof.

25. A method of preventing and/or treating disease associated with a *RPGR* gene comprising administering an agent according to paragraph 23 or a pharmaceutical according to paragraph 24 wherein said agent or said pharmaceutical is capable of modulating a *RPGR* gene or expression product thereof to cause a beneficial therapeutic effect.

26. A kit according to paragraph 3 or paragraph 4 wherein the kit additionally comprises an agent according to paragraph 23 or a pharmaceutical according to paragraph 24; wherein said agent or said pharmaceutical is capable of modulating and/or preventing and/or treating a disease associated with said *RPGR* gene.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be covered by the present invention.

GENERAL TEXT REFERENCES

1. Meindl, A. *et al.* A gene (*RPGR*) with homology to the RCC1 guanine nucleotide exchange factor is mutated in X-linked retinitis pigmentosa (RP3). *Nature Genet.* 13, 35-42 (1996).

2. Roepman, R. *et al.* Positional cloning of the gene for retinitis pigmentosa 3: homology with the guanine-nucleotide-exchange factor RCC1. *Hum. Mol. Genet.* 5, 1035-1041 (1996).

3. Buraczynska, M. *et al.* Spectrum of mutations in the *RPGR* gene that are identified in 20% of families with X-linked retinitis pigmentosa. *Am. J. Hum. Genet.* 61, 1287-92 (1997).

4. Miano, M.G. *et al.* Mutation analysis of the *RPGR* gene reveals novel mutations in South Europian patients with X-linked Retinitis Pigmentosa (RP3) *Eur. J. Hum. Genet.* 7, 687-694 (1999).

5. Ott, J. *et al.* Localizing multiple X chromosome-linked retinitis pigmentosa loci using multilocus homogeneity tests. *Proc. Natl. Acad. Sci. USA.* 87, 701-4 (1990).

6. Teague, P.W. *et al.* Heterogeneity analysis in 40 X-linked retinitis pigmentosa families. *Am. J. Hum. Genet.* 55, 105-11 (1994).

7. Schwahn, U. *et al.* Positional cloning of the gene for X-linked retinitis pigmentosa 2. *Nature Genet.* 19, 327-332 (1998).

8. Hardcastle, A.J. *et al.* Mutations in the *RP2* gene cause disease in 10% of families with familial X-linked retinitis pigmentosa assessed in this study. *Am J. Hum. Genet.* 64, 1210-1215 (1999).

9. Dry, K.L. *et al.* Identification of a novel gene, ETX1 from Xp21.1, a candidate gene for X-linked retintis pigmentosa (RP3). *Hum. Mol. Genet.* 4, 2347-53 (1995).

10. Meindl, A. *et al.* A gene (SRPX) encoding a sushi-repeat-containing protein is deleted in patients with X-linked retinitis pigmentosa. *Hum. Mol. Genet.* 4, 2339-46 (1995).

11. Kirschner, R. *et al.* RPGR transcription studies in mouse and human tissues reveal a retina-specific isoform that is disrupted in a patient with X-linked retinitis pigmentosa. *Hum. Mol. Genet.* 8, 1571-78 (1999).

Wang, G., Seeidman, M.M. & Glazer, P.M. Mutagenesis in mammalian cells induced by triple helix formation and transcription-coupled repair. *Science* 271, 802-805 (1996).

13. Cooper, D.N. & Krawczak, M. *Human Gene Mutation* (Bios, Oxford, 1993).

Weaver, D.T. & DePamphilis, M.L. Specific sequences in native DNA that arrest synthesis by DNA polymerase α. *J. Biol. Chem.* 257, 2075-2086 (1982).

15. Roepman, R. *et al.* Identification of a gene disrupted by a microdeletion in a patient with X-linked retinitis pigmentosa (XLRP). *Hum. Mol. Genet.* 5, 827-33 (1996).

16. Jurka, J., Klonowski, P., Dagman, V. & Pelton, P. CENSOR-a program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.* 20, 119-21 (1996).

17. Zhang, M.Q. Identification of protein coding regions in the human genome by quadratic discriminant analysis. *Proc. Natl. Acad. Sci. USA* 94, 565-8 (1997).

18. Xu, Y., Mural, R., Shah, M. & Uberbacher, E. Recognizing exons in genomic sequence using GRAIL II. *Genet. Eng.* 16, 241-53 (1994).

19. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78-94 (1997).

20. Reese, M.G., Eeckman, F.H., Kulp, D. & Haussler, D. Improved splice site detection in Genie. *J. Comput. Biol.* 4, 311-23 (1997).

21. Williams, G.W., Woollard, P.M. & Hingamp, P. NIX: A Nucleotide Identification system at the HGMP-RC URL: http://www.hgmp.mrc.ac.uk/NIX/ (1998).

22. Dunn, K.C., Aotaki-Keen, A.E., Putkey, F.R. & Hjelmeland, L.M. ARPE-19, a human retinal pigment epithelial cell line with differentiated properties. *Exp. Eye Res.*, 62, 155-69 (1996).

23. Dry, K. Manson, F.D., Lennon, A., Bergen, A.A., Van Dorp, D.B. & Wright, A.F. Identification of a 5' splice site mutation in the RPGR gene in a family with X-linked retinitis pigmentosa (RP3). *Hum. Mutat.* 13, 141-45 (1999)

REFERENCES FROM OMIM SITE

1. Arden, G. B.; Fox, B. : Increased incidence of abnormal nasal cilia in patients with retinitis pigmentosa. Nature 279: 534-536, 1979.

2. Bhattacharya, S. S.; Wright, A. F.; Clayton, J. F.; Price, W. H.; Phillips, C. S.; McKeown, C. M. E.; Jay, M.; Bird, A. C.; Pearson, P. L.; Southern, E. M.; Evans, H. J. : Close genetic linkage between X-linked retinitis pigmentosa and a restriction fragment length polymorphism identified by recombinant DNA probe L1.28. Nature 309: 253-255, 1984.

3. Buraczynska, M.; Wu, W.; Fujita, R.; Buraczynska, K.; Phelps, E.; Andreasson, S.; Bennett, J.; Birch, D. G.; Fishman, G. A.; Hoffman, D. R.; Inana, G.; Jacobson, S. G.; Musarella, M. A.; Sieving, P. A.; Swaroop, A. : Spectrum of mutations in the RPGR gene that are identified in 20% of families with X-linked retinitis pigmentosa. Am. J. Hum. Genet. 61: 1287-1292, 1997.

4. Chen, J. D.; Denton, M. J.; Donald, J.; Serravalle, S.; Colley, P.; Lam Po Tang, P. R. L.; Halliday, F. B. : Analysis of linkage relationships of X-linked retinitis pigmentosa with the following Xp loci: DXS7, OTC, DXS84, XJ-1, DXS164, and DXS28. (Abstract) Cytogenet. Cell Genet. 46: 592, 1987.

5. Chen, J. D.; Dickinson, P.; Gray, R.; Constable, I.; Sheffield, L.; Denton, M. J. : Non-allelic mutations in X-linked retinitis pigmentosa. Clin. Genet. 35: 338-342, 1989. PubMed ID : 2569369 6. Curtis, D.; Blank, C. E. : Recombination between DXS7, DXS84 and a rare form of X-linked retinitis pigmentosa (McK-30320). (Letter) Hum. Genet. 81: 200-202, 1989.

7. Dahl, N.; Sundvall, M.; Pettersson, U.; Andreasson, S.; Anvret, M.; Kugelberg, U.; Hagbyhn-Gericke, A.; Goonewardena, P. : Genetic mapping of loci for X-linked retinitis pigmentosa. Clin. Genet. 40: 435-440, 1991.

8. Denton, M. J.; Chen, J.-D.; Serravalle, S.; Colley, P.; Halliday, F. B.; Donald, J. : Analysis of linkage relationships of X-linked retinitis pigmentosa with the following Xp loci: L1.28, OTC, 754, XJ-1.1, pERT87, and C7. Hum. Genet. 78: 60-64, 1988.

9. Falls, H. F.; Cotterman, C. W. : Choroidoretinal degeneration: a sex-linked form in which heterozygous women exhibit a tapetal-like retinal reflex. Arch. Ophthal. 40: 685-703, 1948.

10. Fishman, G. A.; Grover, S.; Buraczynska, M.; Wu, W.; Swaroop, A. : A new 2-base pair deletion in the RPGR gene in a black family with X-linked retinitis pigmentosa. Arch. Ophthal. 116: 213-218, 1998.

11. Fishman, G. A.; Grover, S.; Jacobson, S. G.; Alexander, K. R.; Derlacki, D. J.; Wu, W.; Buraczynska, M.; Swaroop, A. : X-linked retinitis pigmentosa in two families with a missense mutation in the RPGR gene and putative change of glycine to valine at codon 60. Ophthalmology 105: 2286-2296, 1998.

12. Fox, B.; Bull, T. B.; Arden, G. B. : Variations in the ultrastructure of human nasal cilia including abnormalities found in retinitis pigmentosa. J. Clin. Path. 33: 327-335, 1980.

13. Francke, U.; Ochs, H. D.; de Martinville, B.; Giacalone, J.; Lindgren, V.; Disteche, C.; Pagon, R. A.; Hofker, M. H.; van Ommen, G. J. B.; Pearson, P. L.; Wedgwood, R. J. : Minor Xp21 chromosome deletion in a male associated with expression of Duchenne muscular dystrophy, chronic granulomatous disease, retinitis pigmentosa and McLeod syndrome. Am. J. Hum. Genet. 37: 250-267, 1985.

14. Fujita, R.; Bingham, E.; Forsythe, P.; McHenry, C.; Aita, V.; Vavia, B. A.; Dry, K.; Segal, M.; Devoto, M.; Bruns, G.; Wright, A. F.; Ott, J.; Sieving, P. A.; Swaroop, A. : A recombination outside the BB deletion refines the location of the X-linked retinitis pigmentosa locus RP3. Am. J. Hum. Genet. 59: 152-158, 1996.

15. Fujita, R.; Buraczynska, M.; Gieser, L.; Wu, W.; Forsythe, P.; Abrahamson, M.; Jacobson, S. G.; Sieving, P. A.; Andreasson, S.; Swaroop, A. : Analysis of the RPGR gene in 11 pedigrees with the retinitis pigmentosa type 3 genotype: paucity of mutations in the coding region but splice defects in two families. Am. J. Hum. Genet. 61: 571-580, 1997.

16. Hunter, D. G.; Fishman, G. A.; Kretzer, F. L. : Abnormal axonemes in X-linked retinitis pigmentosa. Arch. Ophthal. 106: 362-368, 1988.

17. Hunter, D. G.; Fishman, G. A.; Mehta, R. S.; Kretzer, F. L. : Abnormal sperm and photoreceptor axonemes in Usher's syndrome. Arch. Ophthal. 104: 385-389, 1986.

18. Keith, C. G.; Denton, M. J.; Chen, J.-D. : Clinical variability in a family with X-linked retinal dystrophy and the locus at the RP3 site. Ophthal. Paediat. Genet. 12: 91-98, 1991.

19. Kirschner, R.; Rosenberg, T.; Schultz-Heienbrok, R.; Lenzner, S.; Feil, S.; Roepman, R.; Cremers, F. P. M.; Ropers, H.-H.; Berger, W. : RPGR transcription studies in mouse and human tissues reveal a retina-specific isoform that is disrupted in a patient with X-linked retinitis pigmentosa. Hum. Molec. Genet. 8: 1571-1578, 1999.

20. Meindl, A.; Dry, K.; Herrmann, K.; Manson, F.; Ciccodicola, A.; Edgar, A.; Carvalho, M. R. S.; Achatz, H.; Hellebrand, H.; Lennon, A.; Migliaccio, C.; Porter, K.; Zrenner, E.; Bird, A.; Jay, M.; Lorenz, B.; Wittwer, B.; D'Urso, M.; Meitinger, T.; Wright, A. : A gene (RPGR) with homology to the RCC1 guanine nucleotide

exchange factor is mutated in X-linked retinitis pigmentosa (RP3). Nature Genet. 13: 35-42, 1996.

21. Miano, M. G.; Testa, F.; Strazzullo, M.; Trujillo, M.; De Bernardo, C.; Grammatico, B.; Simonelli, F.; Mangino, M.; Torrente, I.; Ruberto, G.; Beneyto, M.; Antinolo, G.; Rinaldi, E.; Danesino, C.; Ventruto, V.; D'Urso, M.; Ayuso, C.; Baiget, M.; Ciccodicola, A. : Mutation analysis of the RPGR gene reveals novel mutations in south European patients with X-linked retinitis pigmentosa. Europ. J. Hum. Genet. 7: 687-694, 1999.

22. Musarella, M. A.; Anson-Cartwright, L.; Burghes, A.; Worton, R. G.; Lesko, J. G.; Nussbaum, R. L. : Linkage analysis of a large Latin-American family with X-linked retinitis pigmentosa and metallic sheen in the heterozygote carrier. Genomics 4: 601-605, 1989.

23. Musarella, M. A.; Anson-Cartwright, L.; Leal, S. M.; Gilbert, L. D.; Worton, R. G.; Fishman, G. A.; Ott, J. : Multipoint linkage analysis and heterogeneity testing in 20 X-linked retinitis pigmentosa families. Genomics 8: 286-296, 1990.

24. Musarella, M. A.; Argonza, R.; Burghes, A.; Worton, R. : Location of the gene for XLRP by linkage analysis. (Abstract) Am. J. Hum. Genet. 41: A104, 1987.

25. Nussbaum, R. L.; Lewis, R. A.; Lesko, J. G.; Ferrell, R. : Mapping X-linked ophthalmic diseases: II. Linkage relationship of X-linked retinitis pigmentosa to X chromosomal short arm markers. Hum. Genet. 70: 45-50, 1985.

26. Ott, J.; Bhattacharya, S.; Chen, J. D.; Denton, M. J.; Donald, J.; Dubay, C.; Farrar, G. J.; Fishman, G. A.; Frey, D.; Gal, A.; Humphries, P.; Jay, B.; Jay, M.; Litt, M.; Machler, M.; Musarella, M.; Neugebauer, M.; Nussbaum, R. L.; Terwilliger, J. D.; Weleber, R. G.; Wirth, B.; Wong, F.; Worton, R. G.; Wright, A. F. : Localizing multiple X chromosome-linked retinitis pigmentosa loci using multilocus homogeneity tests. Proc. Nat. Acad. Sci. 87: 701-704, 1990.

27. Roepman, R.; van Duijnhoven, G.; Rosenberg, T.; Pinckers, A. J. L. G.; Bleeker-Wagemakers, L. M.; Bergen, A. A. B.; Post, J.; Beck, A.; Reinhardt, R.; Ropers, H.-H.; Cremers, F. P. M.; Berger, W. : Positional cloning of the gene for X-linked retinitis pigmentosa 3: homology with the guanine-nucleotide-exchange factor RCC1. Hum. Molec. Genet. 5: 1035-1041, 1996.

28. Souied, E.; Segues, B.; Ghazi, I.; Rozet, J.-M.; Chatelin, S.; Gerber, S.; Perrault, I.; Michel-Awad, A.; Briard, M.-L.; Plessis, G.; Dufier, J.-L.; Munnich, A.; Kaplan, J. : Severe manifestations in carrier females in X linked retinitis pigmentosa. J. Med. Genet. 34: 793-797, 1997.

29. van Dorp, D. B.; Wright, A. F.; Carothers, A. D.; Bleeker-Wagemakers, E. M. : A family with RP3 type of X-linked retinitis pigmentosa: an association with ciliary abnormalities. Hum. Genet. 88: 331-334, 1992.

30. Wirth, B.; Bleeker-Wagemakers, E. M.; van Schooneveld, M.; Orth, U.; Neugebauer, M.; Hofker, M.; Pearson, P.; Warburg, M.; Ropers, H. H.; Gal, A. : The gene of X-linked retinitis pigmentosa is closely linked to DXS84 and OTC. (Abstract) Cytogenet. Cell Genet. 46: 718, 1987.

31. Wirth, B.; Denton, M. J.; Chen, J.-D.; Neugebauer, M.; Halliday, F. B.; van Schooneveld, M.; Donald, J.; Bleeker-Wagemakers, E. M.; Pearson, P. L.; Gal, A. : Two different genes for X-linked retinitis pigmentosa. Genomics 2: 263-266, 1988.

CLAIMS

1. A method of diagnosis for a disease or a predisposition to a disease associated with a disease causing mutation(s) in a *RPGR* gene;

wherein the method comprises:

genotyping a *RPGR* gene; and

determining whether the genotype comprises a disease causing mutation(s);

wherein said risk genotype is present within ORF15 of the *RPGR* gene.

2. A method according to claim 1 wherein said risk genotype is present within SEQ ID No. 2.

3. A kit for the diagnosis of a disease or a predisposition to disease; wherein the kit comprises:

means for genotyping a *RPGR* gene; and

reference means for determining whether the genotype comprises a disease causing mutation(s);

wherein said disease causing mutation(s) is present within ORF15 of the *RPGR* gene.

4. A kit according to claim 3 wherein said disease causing mutation(s) is present within SEQ ID No. 2.

5. A nucleotide sequence comprising SEQ ID No. 1 or a variant, homologue, derivative or fragment thereof, but wherein said nucleotide sequence is not, or is not present within, the wild-type *RPGR* gene.

6. A nucleotide sequence comprising SEQ ID No. 2 or a variant, homologue, derivative or fragment thereof, but wherein said nucleotide sequence is not present within the wild-type *RPGR* gene.

7. An amino acid sequence encodable by the sequence of claim 5 or claim 6.

8. A mutant *RPGR* gene, wherein said gene has a mutation in at least ORF15.

9. ORF15 of the *RPGR* gene or the expression product thereof.

10. A nucleotide capable of selectively hybridising to a mutant RPGR gene and not the wild-type *RPGR* gene, wherein said gene has a mutation in at least ORF15, and preferably has one or more disease causing mutation(s).

11. A mutant RPGR protein, wherein said protein has a mutation as a result of a mutation in at least ORF15 of the *RPGR* gene, and preferably has one or more disease causing mutation(s).

12. A kit according to claim 3 or claim 4 wherein the kit additionally comprises an agent capable of modulating and/or preventing and/or treating a disease associated with said disease causing mutation(s) of said *RPGR* gene.

SEQUENCE LISTINGS

SEQ ID NO. 1

ORF15

Both nucleotide and amino acid sequences presented

```
AGATCCCAGAGGAGAAGGAAGGAGCAGAGGATTCAAAAGGAAATGGAATAGAGGAGCAAG 60
                I P E E K E G A E D S K G N G I E E Q E

AGGTAGAAGCAAATGAGGAAAATGTGAAGGTGCATGGAGGAAGAAAGGAGAAAACAGAGA 120
  V E A N E E N V K V H G G R K E K T E I

TCCTATCAGATGACCTTACAGACAAAGCAGAGGTGAGTGAAGGCAAGGCAAAATCAGTGG 180
  L S D D L T D K A E V S E G K A K S V G

GAGAAGCAGAGGATGGGCCTGAAGGTAGAGGGGATGGAACCTGTGAGGAAGGTAGTTCAG 240
  E A E D G P E G R G D G T C E E G S S G

GAGCAGAACACTGGCAAGATGAGGAGAGGGAGAAGGGGGAGAAAGACAAGGGTAGAGGAG 300
  A E H W Q D E E R E K G E K D K G R G E

AAATGGAGAGGCCAGGAGAGGGAGAGAAGGAACTAGCAGAGAAGGAAGAATGGAAGAAGA 360
  M E R P G E G E K E L A E K E E W K K R

GGGATGGGGAAGAGCAGGAGCAAAAGGAGAGGGAGCAGGGCCATCAGAAGGAAAGAAACC 420
  D G E E Q E Q K E R E Q G H Q K E R N Q

AAGAGATGGAGGAGGGAGGGGAGGAGGAGCATGGAGAAGGAGAAGAAGAGGAGGGAGACA 480
  E M E E G G E E E H G E G E E E E G D R

GAGAAGAGGAAGAAGAGAAGGAGGGAGAAGGGAAAGAGGAAGGAGAAGGGGAAGAAGTGG 540
  E E E E E K E G E G K E E G E G E E V E

AGGGAGAACGTGAAAAGGAGGAAGGAGAGAGGAAAAAGGAGGAAAGAGCGGGGAAGGAGG 600
  G E R E K E E G E R K K E E R A G K E E

AGAAAGGAGAGGAAGAAGGAGACCAAGGAGAGGGGGAAGAGGAGGAAACAGAGGGGAGAG 660
  K G E E E G D Q G E G E E E E T E G R G

GGGAGGAAAAAGAGGAGGGAGGGGAAGTAGAGGGAGGGGAAGTAGAGGAGGGGAAAGGAG 720
  E E K E E G G E V E G G E V E E G K G E

AGAGGGAAGAGGAAGAGGAGGAGGGTGAGGGGGAAGAGGAGGAAGGGGAGGGGGAAGAGG 780
  R E E E E E E G E G E E E E G E G E E E

AGGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAAGAAGGGGAAGAAG 840
  E G E G E E E E E G E G K G E E E G E E G

GAGAAGGGGAGGAAGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAG 900
  E G E E E G E E G E G E G E E E E G E G

GGGAGGGAGAAGAGGAAGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGGAGGGAG 960
  E G E E E G E G E G E E E E G E G E G E

AAGAGGAAGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAGGAAG 1020
  E E G E G E G E E E E G E G K G E E E G

GAGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGGAAGGGGAGGATGGAG 1080
  E E G E G E G E E E E G E G E G E D G E

AAGGGGAGGGGGAAGAGGAGGAAGGAGAATGGGAGGGGGAAGAGGAGGAAGGAGAAGGGG 1140
```

```
  G  E  G  E  E  E  E  G  E  W  E  G  E  E  E  E  G  E  G  E

AGGGGGAAGAGGAAGGAGAAGGGGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGG 1200
  G  E  E  E  G  E  G  E  G  E  E  G  E  G  E  G  E  E  E  E

AAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGGGAAGAAGAAGGGGAGGAAGAAGGAGAGG 1260
  G  E  G  E  G  E  E  E  E  G  E  E  E  G  E  E  E  G  E  G

GAGAGGAAGAAGGGGAGGGAGAAGGGGAGGAAGAAGAGGAAGGGGAAGTGGAAGGGGAGG 1320
  E  E  E  G  E  G  E  G  E  E  E  E  E  G  E  V  E  G  E  V

TGGAAGGGGAGGAAGGAGAGGGGGAAGGAGAGGAAGAGGAAGGAGAGGAGGAAGGAGAAG 1380
  E  G  E  E  G  E  G  E  G  E  E  E  E  G  E  E  E  G  E  E

AAAGGGAAAAGGAGGGGGAAGGAGAAGAAAACAGGAGGAACAGAGAAGAGGAGGAGGAAG 1440
  R  E  K  E  G  E  G  E  E  N  R  R  N  R  E  E  E  E  E  E

AAGAGGGGAAGTATCAGGAGACAGGCGAAGAAGAGAATGAAAGGCAGGATGGAGAGGAGT 1500
  E  G  K  Y  Q  E  T  G  E  E  E  N  E  R  Q  D  G  E  E  Y

ACAAAAAAGTGAGCAAAATAAAAGGATCTGTGAAATATGGCAAACATAAAACATATCAAA 1560
  K  K  V  S  K  I  K  G  S  V  K  Y  G  K  H  K  T  Y  Q  K

AAAAGTCAGTTACTAACACACAGGGAAATGGGAAAGAGCAGAGGTCCAAAATGCCAGTCC 1620
  K  S  V  T  N  T  Q  G  N  G  K  E  Q  R  S  K  M  P  V  Q

AGTCAAAACGACTTTTAAAAAATGGGCCATCAGGTTCCAAAAAGTTCTGGAATAATATAT 1680
  S  K  R  L  L  K  N  G  P  S  G  S  K  K  F  W  N  N  I  L

TACCACATTACTTGGAATTGAAGTAACAAACCTTAAATGTGACCCGATTATGGCCAGTCA 1740
  P  H  Y  L  E  L  K  *

GACAATTTAAATGCCTTGCATATAACGGGCACTCATTACGTGTTATTAAATTGATTTTAT 1800
GTCAATTATTTTATGTGTAGTAAAAAAAAAGCAACTGATGCAGCTGTGTTAAGGAGCCAA 1860
AGACAATAGGAGGCACTGGTAAATTTTGGCCTCTCTCAAACTAAAATTTTCGTGTATTTC 1920
CCCCCCAAATTATAAAAACATAACTAGAAAATATTAAAAGGTCATATCAGATTATTAACA 1980
TTATATATTCATTAAAGGCAGCTTTAGGAAACAGGAATATACTACAAGAGTGTTTTGTTT 2040
GTGTATACAAATCATTCCATTTTTAAATGGCACAGATGCTTAAGGGCTATAAAAACTTCT 2100
AATTTCTTATAAATATGTTAGCACTTTTTTTAAGTTAGTGATTACAGTTTACCTACTGTA 2160
TAGAATAATTTTCTAATAATGGATGGTATTCTAAAACTCAATTGAGGCATTCACATTTTA 2220
AAGGAAGTATTGTCTTTCACCTTTTATGTGTTCTTTTTGCAAAAATCTACAAAGTGACAG 2280
CTGTGTTCAGAGCTTAGATCCCAAAAACGTGATCTCTTTTAGTTACTATCTGGGCAGATG 2340
GTAGTATATCTAATGAAATGGTGATTAATTTAAATGTATAATCTGGAAATATGTAAAACT 2400
TGAAGTATTTTTTGTCCAGGCAAAGGTACTCATTGGCCTCAGTTCTCCATCTCTAAAATG 2460
GAGTGGATGAGATGATGTGATAACTGCAGTCCCTTCTAACTCTTAAATTCTTTTCATTCT 2520
CACAGATTCACTCTATCATTATTGTTATTCATGTAAGAAACGTTTTAGGGAGAAAAATTA 2580
CACTTTAAAATTAATTTAGTTTTCTATACAGTTGTTTTCTTTACTCTTGAAAAGTTATGA 2640
CAGCTTTAACGTCTCTTGTCTTCTGTAATTTTTTATTTCTAAACTCCTATCATTTCCAAG 2700
CTTTAACTGTACTTTATCAGAGCCTTCATTTCTGGTATGTGTTATATGCCCTCAATGTAT 2760
TCACTGACTGTTCTGTAATTTCAGTTTGTCTGTTCCTTGTCAGAATGTTTCAAGTAAAAT 2820
AAAAAATTAAATGTAAAAAAAAAAAAAA                                   2834
```

SEQ ID No. 2
MUTATIONAL HOT SPOT

Nucleotide sequence presented (for the corresponding amino acid sequence, see SEQ ID No. 1)

GGGAGCAGGGCCATCAGAAGGAAAGAAACCAAGAGATGGAGGAGGGAGGGGAGGAGGAGC
ATGGAGAAGGAGAAGAAGAGGAGGGAGACA
GACAAGAGGAAGAAGAGAAGGAGGGAGAAGGGAAAGAGGAAGGAGAAGGGGAAGAAGTGG
AGGGAGAACGTGAAAAGGAGGAAGGAGAGAGGAAAAAGGAGGAAAGAGCGGGGAAGGAGG
AGAAAGGAGAGGAAGAAGGAGACCAAGGAGAGAGGGGGAAGAGGAGGAAACAGAGGGAGAG
GGGAGGAAAAAGAGGAGGGAGGGGAAGTAGACGGAGGGGAAGTAGAGGAGGGGAAAGGAG
AGAGGGAAGAGGAAGAGGAGGAGGGTGAGGGGGAAGAGGAGGAAGGGGAGGGGGAAGAGG
AGGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAAGAAGGGGAAGAAG
GAGAAGGGGAGGAAGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAG
GGGAGGGAGAAGAGGAAGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGGAGGGAG
AAGAGGAAGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAAG
GAGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGGAAGGGGAGGATGGAG
AAGGGGAGGGGGAAGAGGAGGAAGGAGAATGGGAGGGGGAAGAGGAGGAAGGAGAAGGGG
AGGGGGAAGAGGAAGGAGAAGGGGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGG
AAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGGGAAGAAGAAGGGGAGGAAGAAGGAGAGG
GAGAGGAAGAAGGGGAGGGAGAAGGGGAGGAAGAAGAGGAAGGGGAAGTGGAAGGGGAGG
TGGAAGGGGAGGAAGGAGAGGGGCAAGGAGAGGAAGAGGAAGGAGAGGAGGAAGGAGAAG
AAAGGGAAAAGGAGGGGGAAGGAGAAGAAAACAGGAGGAACAGAGAAGAGGAGGAGGAAG

SEQ ID No. 3

This specific sequence listing (SEQ ID No. 3) covers one or more of the sequences presented on a dark background

Nucleotide sequences are presented (for the corresponding amino acid sequences, see SEQ ID No. 1)

GGGAGCAGGGCCATCAGAAGGAAAGAAACCAAGAGATGGAGGAGGGAGGGGAGGAGGAGC
ATGGAGAAGGAGAAGAAGAGGAGGGAGACA
GACAAGAGGAAGAAGAGAAGGAGGGAGAAGGGAAAGAGGAAGGAGAAGGGGAAGAAGTGG
AGGGAGAACGTGAAAAGGAGGAAGGAGAGAGGAAAAAGGAGGAAAGAGCGGGGAAGGAGG
AGAAAGGAGAGGAAGAAGGAGACCAAGGAGAGGGGGAAGAGGAGGAAACAGAGGGAGAG
GGGAGGAAAAAGAGGAGGGAGGGGAAGTAGAGGGGAAGTAGAGGAGGGGAAAGGAG
AGAGGGAAGAGGAAGAGGAGGAGGGTGAGGGGGAAGAGGAGGAAGGGGAGGGGAAGAGG
AGGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAAGAAGGGGAAGAAG
GAGAAGGGGAGGAAGAAGGGGAGGAAGGAGAAGGGGAGGGGAAGAGGAGGAAGGAGAAG
GGGAGGGAGAAGAGGAAGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGGAGGGAG
AAGAGGAAGGAGAAGGGGAGGGGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAAGAAG
GAGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGGAAGGGGAGGATGGAG
AAGGGGAGGGGGAAGAGGAGGAAGGAGAATGGGAGGGGGAAGAGGAGGAAGGAGAAGGGG
AGGGGGAAGAGGAAGGAGAAGGGGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGG
AAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGGGAAGAAGAAGGCGGAGGAAGAAGGAGAGG
GAGAGGAAGAAGGGGAGGGAGAAGGGGAGGAAGAAGAAGAGGAAGGGGAAGTGGAAGGGGAGG
TGGAAGGGGAGGAAGGAGAGGGGCAAGGAGAGGAAGAGGAAGGAGAGGAGGAAGGAGAAG
AAAGGGAAAAGGAGGGGGAAGGAGAAGAAAACAGGAGGAACAGAGAAGAGGAGGAGGAAG

FIG. 1

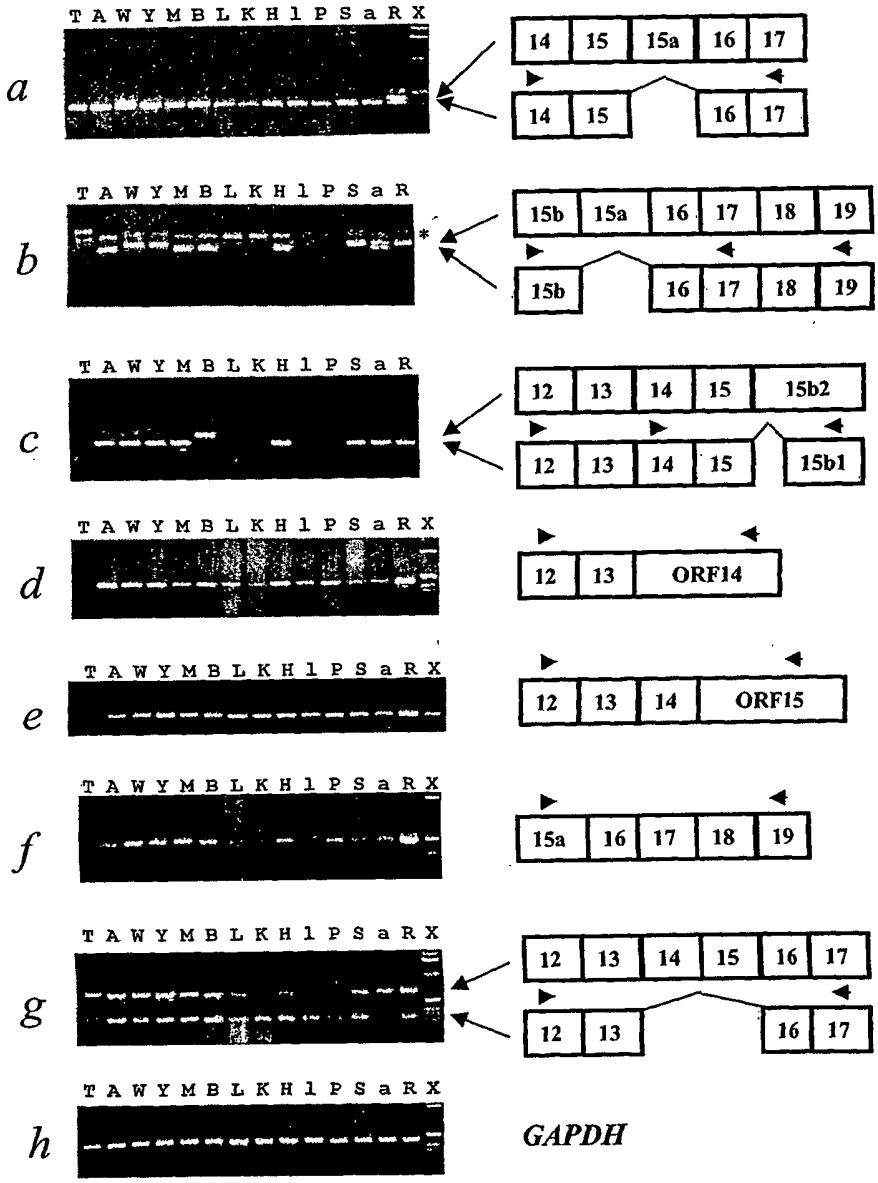
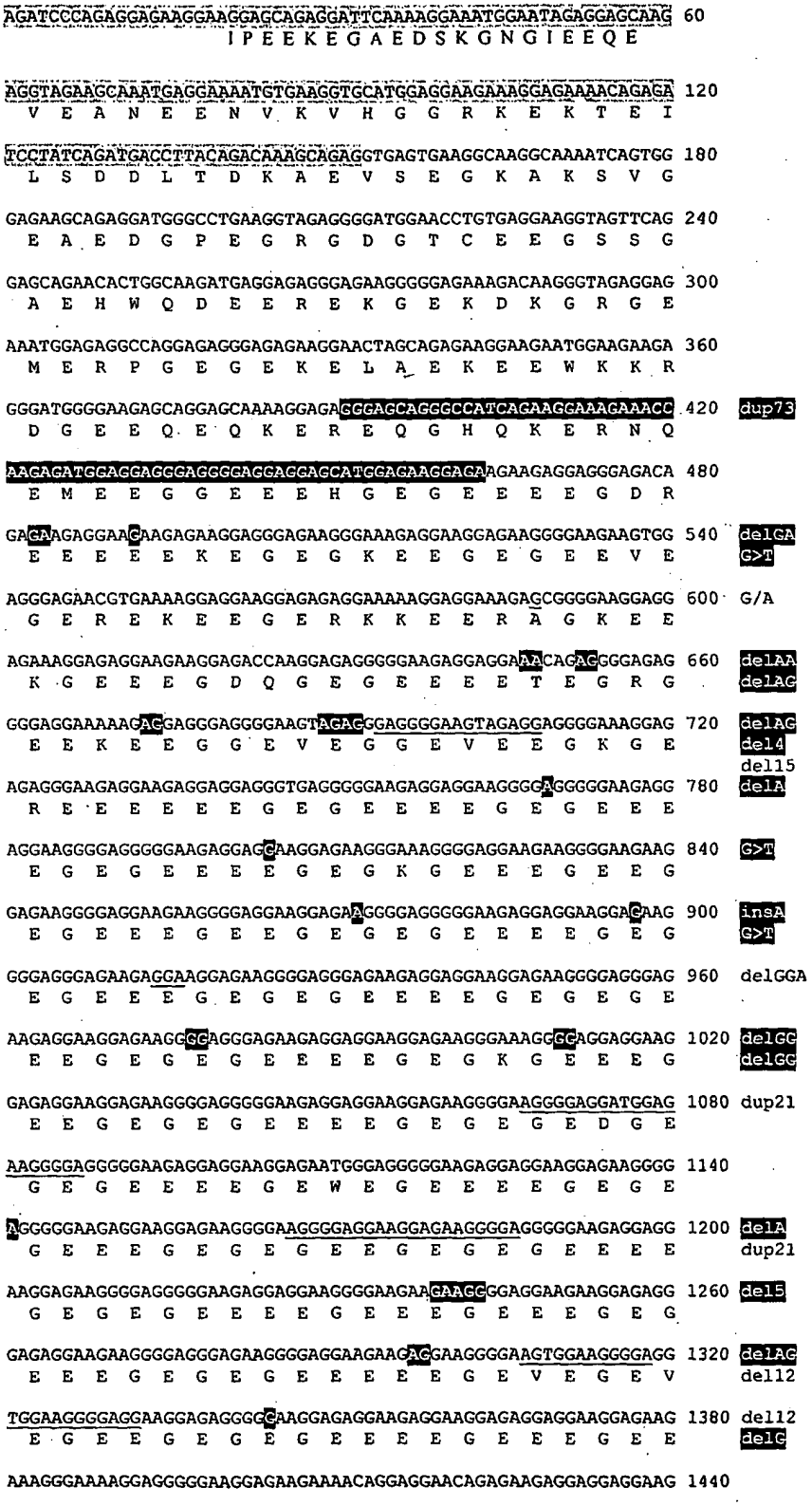

T A W Y M B L K H 1 P S a R X
a
14 15 15a 16 17
14 15 16 17
T A W Y M B L K H 1 P S a R
b
*
15b 15a 16 17 18 19
15b 16 17 18 19
T A W Y M B L K H 1 P S a R
c
12 13 14 15 15b2
12 13 14 15 15b1
T A W Y M B L K H 1 P S a R X
d
12 13 ORF14
T A W Y M B L K H 1 P S a R X
e
12 13 14 ORF15
T A W Y M B L K H 1 P S a R X
f
15a 16 17 18 19
T A W Y M B L K H 1 P S a R X
g
12 13 14 15 16 17
12 13 16 17
T A W Y M B L K H 1 P S a R X
h
GAPDH

FIG. 2

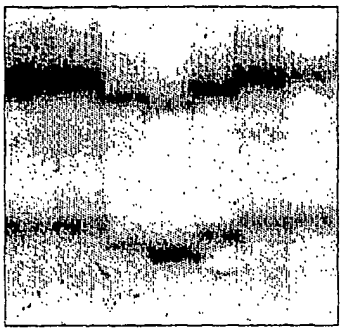

AGATCCCAGAGGAGAAGGAAGGAGCAGAGGATTCAAAAGGAAATGGAATAGAGGAGCAAG 60
I P E E K E G A E D S K G N G I E E Q E
AGGTAGAAGCAAATGAGGAAAATGTGAAGGTGCATGGAGGAAGAAAGGAGAAAACAGAGA 120
V E A N E E N V K V H G G R K E K T E I
TCCTATCAGATGACCTTACAGACAAAGCAGAGGTGAGTGAAGGCAAGGCAAAATCAGTGG 180
L S D D L T D K A E V S E G K A K S V G
GAGAAGCAGAGGATGGGCCTGAAGGTAGAGGGGATGGAACCTGTGAGGAAGGTAGTTCAG 240
E A E D G P E G R G D G T C E E G S S G
GAGCAGAACACTGGCAAGATGAGGAGAGGGAGAAGGGGGAGAAAGACAAGGGTAGAGGAG 300
A E H W Q D E E R E K G E K D K G R G E
AAATGGAGAGGCCAGGAGAGGGAGAGAAGGAACTAGCAGAGAAGGAAGAATGGAAGAAGA 360
M E R P G E G E K E L A E K E E W K K R
GGGATGGGGAAGAGCAGGAGCAAAAGGAGAGGGAGCAGGGCCATCAGAAGGAAAGAAACC 420 dup73
D G E E Q E Q K E R E Q G H Q K E R N Q
AAGAGATGGAGGAGGGAGGGGAGGAGGAGCATGGAGAAGGAGAAGAGGAGGGAGACA 480
E M E E G G E E E H G E G E E E E G D R
GAGAAGAGGAAGAAGAGAAGGAGGGAGAAGGGAAAGAGGAAGGAGAAGGGGAAGAAGTGG 540 delGA
E E E E E K E G E G K E E G E G E E V E G>T
AGGGAGAACGTGAAAAGGAGGAAGGAGAGAGGAAAAAGGAGGAAAGAGCGGGGAAGGAGG 600 G/A
G E R E K E E G E R K K E E R A G K E E
AGAAAGGAGAGGAAGAAGGAGACCAAGGAGAGGGGGAAGAGGAGGAAACAGAGGGGAGAG 660 delAA
K G E E E G D Q G E G E E E E T E G R G delAG
GGGAGGAAAAAGAGGAGGGAGGGGAAGTAGACGGAGGGGAAGTAGAGGAGGGGAAAGGAG 720 delAG
E E K E E G G E V E G G E V E E G K G E del4
del15
AGAGGGAAGAGGAAGAGGAGGAGGGTGAGGGGGAAGAGGAGGAAGGGGAGGGGGAAGAGG 780 delA
R E E E E E E G E G E E E E G E G E E E
AGGAAGGGGAGGGGGAAGAGGAGCAAGGAGAAGGGAAAGGGGAGGAAGAAGGGGAAGAAG 840 G>T
E G E G E E E E E G E G K G E E E G E E G
GAGAAGGGGAGGAAGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAG 900 insA
E G E E E E G E E G E G E G E E E E E G E G G>T
GGGAGGGAGAAGAGGAAGGAGAAGGGGAGGGAGAAGAGGAGGAAGGAGAAGGGGAGGGAG 960 delGGA
E G E E E G E G E G E E E E G E G E G E
AAGAGGAAGGAGAAGGGGAGGGAGGAAGAGGAGGAAGGAGAAGGGAAAGGGGAGGAAG 1020 delGG
E E G E G E G E E E E G E G K G E E E G delGG
GAGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGAGAAGGGGAAGGGGAGGATGGAG 1080 dup21
E E G E G E G E E E E G E G E G E D G E
AAGGGGAGGGGGAAGAGGAGGAAGGAGAATGGGAGGGGGAAGAGGAGGAAGGAGAAGGGG 1140
G E G E E E E G E W E G E E E E G E G E
AGGGGGAAGAGGAAGGAGAAGGGGAAGGGGAGGAAGGAGAAGGGGAGGGGGAAGAGGAGG 1200 delA
G E E E G E G E G E E G E G E G E E E E dup21
AAGGAGAAGGGGAGGGGGAAGAGGAGGAAGGGGAAGAAGAAGGGGAGGAAGAAGGAGAGG 1260 del5
G E G E G E E E E G E E E G E E E G E G
GAGAGGAAGAAGGGGAGGGAGAAGGGGAGGAAGAAGAGGAAGGGGAAGTGGAAGGGGAGG 1320 delAG
E E E G E G E G E E E E E E G E V E G E V del12
TGGAAGGGGAGGAAGGAGAGGGGCAAGGAGAGGAAGAGGAAGGAGAGGAGGAAGGAGAAG 1380 del12
E G E E G E G E G E E E E G E E E G E E delG
AAAGGGAAAAGGAGGGGGAAGGAGAAGAAAACAGGAGGAACAGAGAAGAGGAGGAGGAAG 1440

```
  R  E  K  E  G  E  G  E  E  N  R  R  N  R  E  E  E  E  E  E

AAGAGGGGAAGTATCAGGAGACAGGCGAAGAAGAGAATGAAAGGCAGGATGGAGAGGAGT 1500 C/T
  E  G  K  Y  Q  E  T  G  E  E  E  N  E  R  Q  D  G  E  E  Y

ACAAAAAAGTGAGCAAAATAAAAGGATCTGTGAAATATGGCAAACATAAAACATATCAAA 1560
  K  K  V  S  K  I  K  G  S  V  K  Y  G  K  H  K  T  Y  Q  K

AAAAGTCAGTTACTAACACACAGGGAAATGGGAAAGAGCAGAGGTCCAAAATGCCAGTCC 1620
  K  S  V  T  N  T  Q  G  N  G  K  E  Q  R  S  K  M  P  V  Q

AGTCAAAACGACTTTTAAAAAATGGGCCATCAGGTTCCAAAAAGTTCTGGAATAATATAT 1680
  S  K  R  L  L  K  N  G  P  S  G  S  K  K  F  W  N  N  I  L

TACCACATTACTTGGAATTGAAGTAACAAACCTTAAATGTGACCCGATTATGGCCAGTCA 1740
  P  H  Y  L  E  L  K  *

GACAATTTAAATGCCTTGCATATAACGGGCACTCATTACGTGTTATTAAATTGATTTTAT 1800
GTCAATTATTTTATGTGTAGTAAAAAAAAAGCAACTGATGCAGCTGTGTTAAGGAGCCAA 1860
AGACAATAGGAGGCACTGGTAAATTTTGGCCTCTCTCAAACTAAAATTTTCGTGTATTTC 1920
CCCCCCAAATTATAAAAACATAACTAGAAAATATTAAAAGGTCATATCAGATTATTAACA 1980
TTATATATTCATTAAAGGCAGCTTTAGGAAACAGGAATATACTACAAGAGTGTTTTGTTT 2040
GTGTATACAAATCATTCCATTTTTAAATGGCACAGATGCTTAAGGGCTATAAAAACTTCT 2100
AATTTCTTATAAATATGTTAGCACTTTTTTTAAGTTAGTGATTACAGTTTACCTACTGTA 2160
TAGAATAATTTTCTAATAATGGATGGTATTCTAAAACTCAATTGAGGCATTCACATTTTA 2220
AAGGAAGTATTGTCTTTCACCTTTTATGTGTTCTTTTTGCAAAAATCTACAAAGTGACAG 2280
CTGTGTTCAGAGCTTAGATCCCAAAAACGTGATCTCTTTTAGTTACTATCTGGGCAGATG 2340
GTAGTATATCTAATGAAATGGTGATTAATTTAAATGTATAATCTGGAAATATGTAAAACT 2400
TGAAGTATTTTTTGTCCAGGCAAAGGTACTCATTGGCCTCAGTTCTCCATCTCTAAAATG 2460
GAGTGGATGAGATGATGTGATAACTGCAGTCCCTTCTAACTCTTAAATTCTTTTCATTCT 2520
CACAGATTCACTCTATCATTATTGTTATTCATGTAAGAAACGTTTTAGGGAGAAAAATTA 2580
CACTTTAAAATTAATTTAGTTTTCTATACAGTTGTTTTCTTTACTCTTGAAAAGTTATGA 2640
CAGCTTTAACGTCTCTTGTCTTCTGTAATTTTTTATTTCTAAACTCCTATCATTTCCAAG 2700
CTTTAACTGTACTTTATCAGAGCCTTCATTTCTGGTATGTGTTATATGCCCTCAATGTAT 2760
TCACTGACTGTTCTGTAATTTCAGTTTGTCTGTTCCTTGTCAGAATGTTTCAAGTAAAAT 2820
AAAAAATTAAATGTAAAAAAAAAAAAAA                                  2834
```

FIG. 2 CONT'D

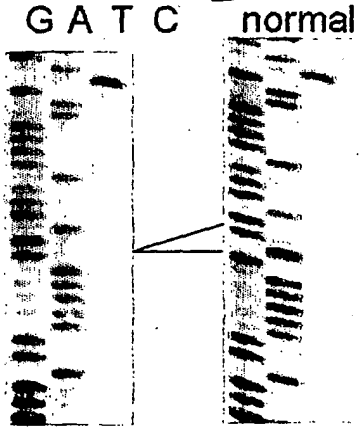
FIG. 3

a

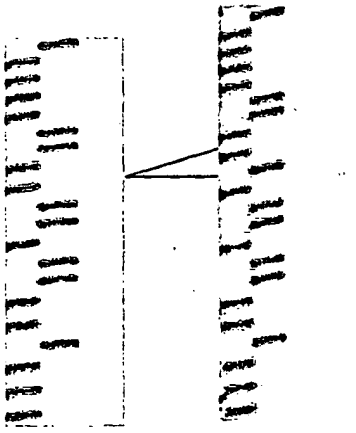


b

g.ORF15+673_674delAG

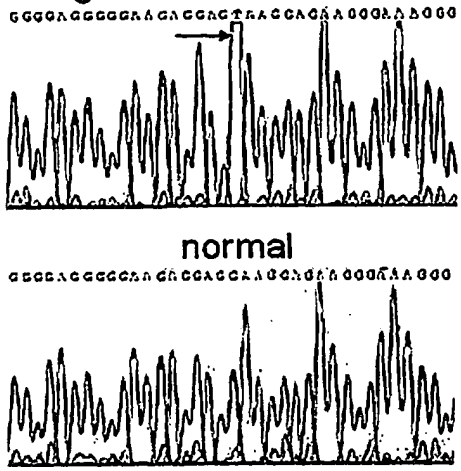


g.ORF15+1297_1298delAG

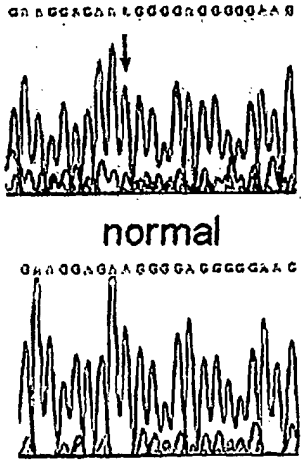


c

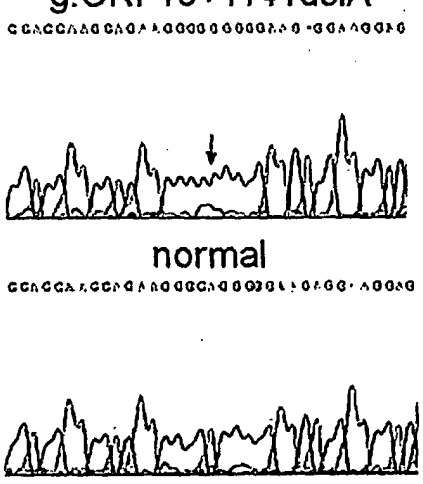
g.ORF15+804G>T



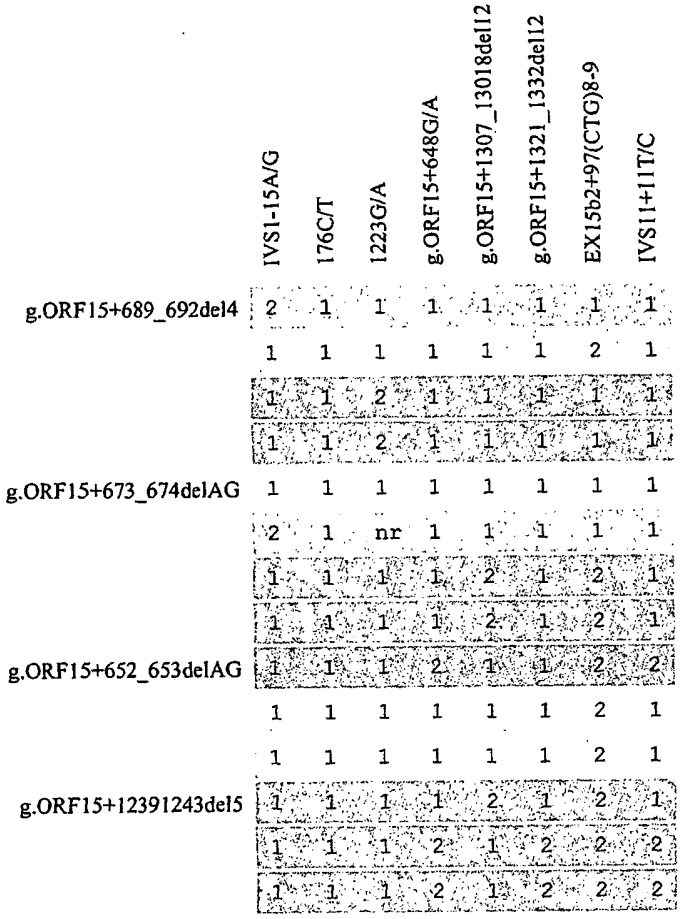
g.ORF15+872_873insA



g.ORF15+1141delA



d



e

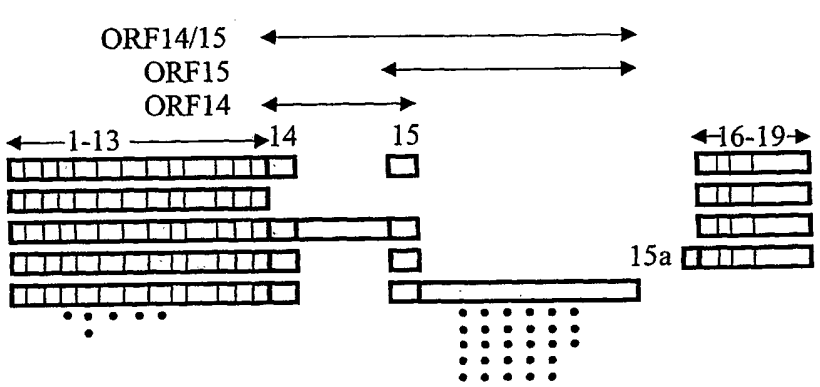
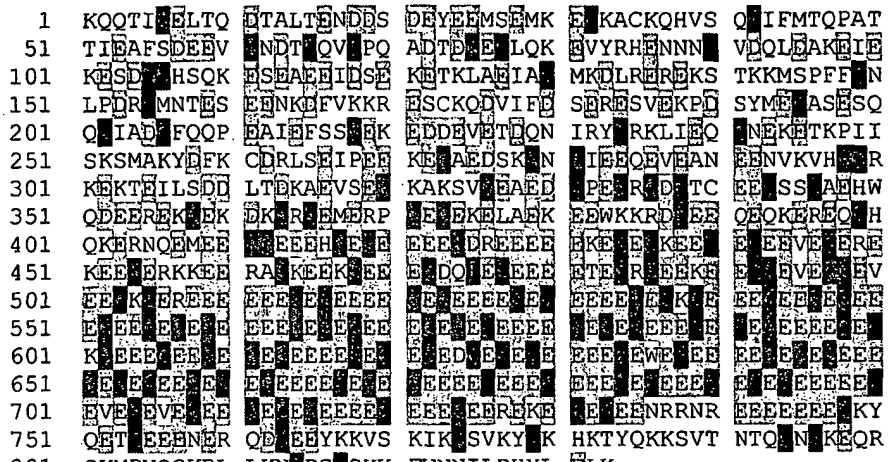


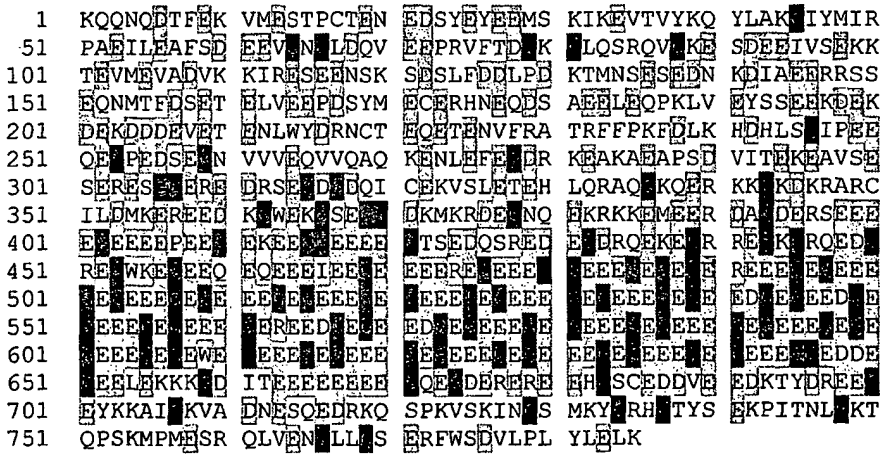
FIG. 3 CONT'D

FIG. 4

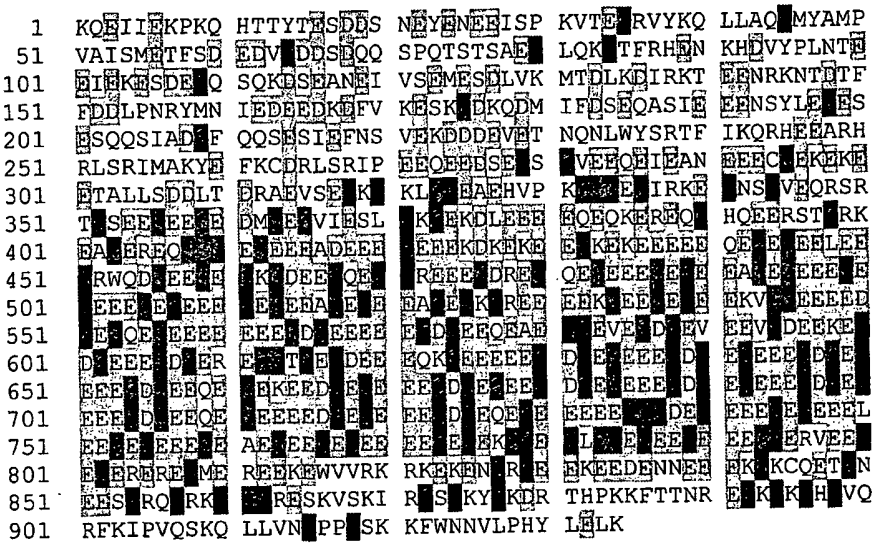
a



b



c



d

FIG. 4 CONT'D

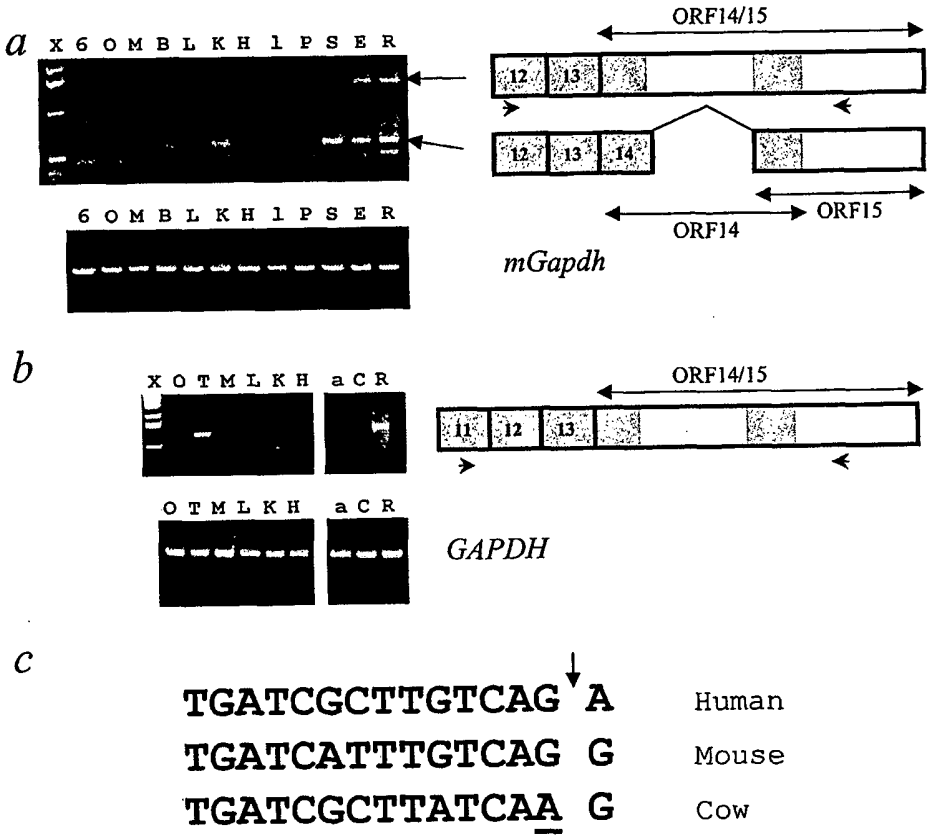
e

f

```
KKSVTNTQGN GK...EQRSK MPVQSKRLLK NGPSGSKKFW NNILPHYLPLK  H
KKFTTNREGK GKGHGVQRFK IPVQSKQLLV NGPPGSKKFW NNVLPHYLPLK  B
EKPITN...L GK...TQPSK MPMESRQLVE NGLLGSERFW SDVLPLYLPLK  M
EEEEIK...K KK....RPVK LGRQHRSGVK GQAAEPKEFW GDVLPQYLNLK  F
```

FIG. 4 CONT'D

FIG. 5


a
X 6 O M B L K H 1 P S E R
6 O M B L K H 1 P S E R
ORF14/15
12
13
14
ORF14
ORF15
mGapdh
b
X O T M L K H aC R
O T M L K H aC R
ORF14/15
11
12
13
GAPDH
c
TGATCGCTTGTCAG A Human
TGATCATTTGTCAG G Mouse
TGATCGCTTATCAA G Cow